

THIS WEEK

EDITORIALS

WORLD VIEW Europe needs a research leader who will lead **p.7**

NEUROSCIENCE Brain activity could help to predict TV success **p.8**



ZOOLOGY Manta rays change colour to communicate **p.9**

Home-brew tests need regulation

A US proposal to regulate medical diagnostics from individual labs reflects the tests' growing complexity. Such guidance should be welcomed, not resisted.

On 18 April, the US Centers for Disease Control and Prevention (CDC) published an alert. The agency had learned of a new test being used to diagnose Lyme disease, a tick-borne bacterial infection that can cause fatigue, joint pain and nervous-system problems. The test, like many others for the disease, had not been formally evaluated and approved by governmental regulators, and agency scientists worried that the method would churn out too many false positives. But because of a regulatory loophole, there was little the CDC could do except ask consumers to avoid the tests and urge people to seek out the few diagnostics that had been approved by the US Food and Drug Administration (FDA).

The problem extends well beyond Lyme disease. Thousands of other 'home-brew' medical tests — those developed in individual laboratories and used to guide the diagnosis and treatment of everything from cancer to *Candida* — have largely escaped federal oversight.

That is now likely to change. On 31 July, the FDA unveiled its plans to regulate the field. In doing so, the agency is risking the wrath of industry and academic labs alike, which have argued that regulation of home-brew tests will slow the development of diagnostics unnecessarily. Yet expanded oversight is warranted, and researchers would do well to learn from the FDA's example. As medical diagnostics become more elaborate and more important in health-care decisions, they need to be treated with more gravity.

In 1976, the US Congress declared that most diagnostic tests could be considered medical devices and therefore fell under the FDA's regulatory purview. But at the time, laboratory tests tended to be simple, familiar assays performed using components that had been approved for clinical use. Typically, physicians and pathologists — often at the same institution that carried out the test — interpreted the results. Given this relatively safe environment, the FDA exercised its discretion and declared that it would not regulate home-brew tests. (The FDA does, however, regulate commercial tests that are developed and then sold as kits to be used in other labs.)

A COMPLEX BREW

Today, the medical-diagnostics field is very different. Tests are used more frequently, and in higher-risk settings, to select therapies for critically ill patients. Although some familiar tests remain, home-brew tests are increasingly carried out using cutting-edge science and technologies, and yield results so complex that they require proprietary algorithms to parse the data. Genome-wide surveys of gene-expression patterns and genomic abnormalities, for example, have emerged as attractive ways to select treatments for people with cancer. But they present challenges for standardization across labs.

The business of laboratory testing has also changed, with many tests now provided by large companies that mass-market their products. The well-known test for cancer-associated mutations in the genes *BRCA1* and *BRCA2*, for example — provided by Myriad Genetics of

Salt Lake City, Utah — is a home-brew assay because its results are not independently analysed outside the company. Although that test has a substantial body of research backing its veracity, many other tests do not. And whereas regulators inspect general techniques and equipment at some of these labs, they generally do not ascertain the validity of the particular tests the labs deploy.

The FDA announced its intentions to change this policy at least as early as 2010. Opposition was swift and fierce, and came from both industry and academia. The long delay in the release of the FDA's new policy prompted rumours of political interference. In July, five US senators wrote a letter to the Office of Management and Budget, which has to review proposed regulations, to question the delay in releasing the FDA's guidance.

But in another letter sent last month, a host of academic testing labs decried efforts to regulate the field, saying that the tests should be considered services rather than devices. It is easy to understand some of their concerns. The FDA is famously overcommitted and under-resourced, and adding to its remit raises fears that the agency will be slow to issue approvals, becoming a roadblock to innovation just as the technologies are beginning to build up speed.

Fortunately, the plans unveiled by the FDA may sidestep such concerns. The regulations will be phased in gradually, to avoid abrupt interruptions of important medical services. And the agency intends to focus first on tests that bear the most risk for patients. Low-risk tests and those for rare diseases are likely to be excluded from regulation.

Properly executed, the proposals could bring welcome scientific rigour to a field that has become unruly. Some FDA staff say that they have struggled to combat the outdated sense of complacency with respect to medical tests — and not only in clinical pathology labs. Researchers, too, have had to be persuaded that diagnostics deserve heightened scrutiny. Too many scientists are still not aware that the agency needs to review trials involving medical tests — for example, clinical trials that select cancer therapies on the basis of mutations found in a participant's tumour. If such a trial is considered sufficiently risky, the FDA may require further evidence that the test is valid.

Researchers sometimes chafe at these rules. But in 2010, Duke University in Durham, North Carolina, ended three clinical trials designed to determine whether gene-expression profiles could predict patient responses to lung-cancer therapies. The trials were based on results from cancer researcher Anil Potti, and were terminated well after other scientists reported flaws in his analyses. Those flaws might have been acknowledged earlier if the FDA had been consulted before the trials started.

With its proposal to regulate home-brew tests, the FDA is responding to a changing medical climate. Researchers must be willing to do the same. ■

"The proposals could bring welcome scientific rigour to a field that has become unruly."

early as 2010. Opposition was swift and fierce, and came from both industry and academia. The long delay in the release of the FDA's new policy prompted rumours of political interference. In July, five US senators wrote a letter to the Office of Management and Budget, which has to review proposed regulations, to question the delay in releasing the FDA's guidance.

But in another letter sent last month, a host of academic testing labs decried efforts to regulate the field, saying that the tests should be considered services rather than devices. It is easy to understand some of their concerns. The FDA is famously overcommitted and under-resourced, and adding to its remit raises fears that the agency will be slow to issue approvals, becoming a roadblock to innovation just as the technologies are beginning to build up speed.

Fortunately, the plans unveiled by the FDA may sidestep such concerns. The regulations will be phased in gradually, to avoid abrupt interruptions of important medical services. And the agency intends to focus first on tests that bear the most risk for patients. Low-risk tests and those for rare diseases are likely to be excluded from regulation.

Properly executed, the proposals could bring welcome scientific rigour to a field that has become unruly. Some FDA staff say that they have struggled to combat the outdated sense of complacency with respect to medical tests — and not only in clinical pathology labs. Researchers, too, have had to be persuaded that diagnostics deserve heightened scrutiny. Too many scientists are still not aware that the agency needs to review trials involving medical tests — for example, clinical trials that select cancer therapies on the basis of mutations found in a participant's tumour. If such a trial is considered sufficiently risky, the FDA may require further evidence that the test is valid.

Researchers sometimes chafe at these rules. But in 2010, Duke University in Durham, North Carolina, ended three clinical trials designed to determine whether gene-expression profiles could predict patient responses to lung-cancer therapies. The trials were based on results from cancer researcher Anil Potti, and were terminated well after other scientists reported flaws in his analyses. Those flaws might have been acknowledged earlier if the FDA had been consulted before the trials started.

With its proposal to regulate home-brew tests, the FDA is responding to a changing medical climate. Researchers must be willing to do the same. ■



Europe needs a research leader who will lead

The next research commissioner for the European Union will need the drive and confidence to clear a daunting in-tray, argues Colin Macilwain.

In a few months, Europe will appoint a research commissioner whose €11-billion (US\$15-billion) annual budget will make him or her, at least in theory, the most influential figure in European science policy.

Nations are now scrambling to pick and send to Brussels one commissioner each, to provide a pool of 28 from which the research head and others will be plucked. Research commissioner is not the most prestigious appointment for some of these people. But it is a crucial one for Europe's researchers, many of whom are spending rather too much time grumpily pondering career prospects — their own, and others' — in the United States or Asia.

The right appointment could help to lift their morale. The wrong one could squander the promise of Horizon 2020, the €80-billion, 7-year research-and-innovation programme that the European Union (EU) instigated this year.

The longlist is far from complete, so it is too early to speculate on who will get the job. What is known is that one of their first tasks will be to get Horizon 2020 firmly back on the rails. Some elements of the programme, notably the European Research Council, are in reassuringly rude health. But there are already ominous murmurs among researchers that Horizon 2020 could fail to deliver on its promise to address 'grand challenges' such as ageing and climate change.

Horizon 2020 relies on an array of old 'instruments' with unsexy names, such as Joint Technology Initiatives, to tackle these challenges. But it is not clear that the mix of instruments has the necessary cohesion to make a visible impact on the challenges. And many talented university researchers, who still live out their lives in disciplinary silos, seem to have baulked at applying for early Horizon 2020 calls that are phrased in terms of those broad, societal goals.

While addressing these problems, the commissioner will have to calm ongoing turmoil in the administration of the research directorate itself. Hundreds of staff members who deal with research proposals are being dispersed to agencies outside the commission. They are unlikely to go quietly. Such extensive reorganization tends — at least in the short term — to trigger turf wars and backbiting that lower morale and clog the system.

The default position of the staff involved, as with most civil servants, is to loftily declare that they expect little — and receive less — in the way of support or inspiration from their boss, the commissioner. But in real life, leadership does matter. Only a strong and visible commissioner will allow the directorate to operate effectively while these administrative changes take place.

The commissioner must also have long-term

vision. That sounds like a cliché, but it happens to be true. The EU's unique seven-year budget process means that the plans for after 2020 need to be developed on the new commissioner's watch.

For example, Horizon 2020 was built around three basic ideas — the grand challenges, more emphasis on innovation and a larger European Research Council — that were firmly in place years ago. That outline had taken shape before the financial crisis struck in 2008, hammering national budgets and leaving researchers in swathes of Europe with minimal funding or job prospects.

The crisis should have triggered a rethink on how research money and other funds could be used to shore up opportunities in regions of eastern and southern Europe where the research base is crumbling.

But despite some late window-dressing, Horizon 2020 doesn't really take this issue seriously.

To address this, the commission is undertaking a consultation on 'Science 2.0', the buzzword for its vision of how science should be done and organized. What is a peer-reviewed paper? Whose data is it based on? Who are its authors? As Europe's largest research funder, the commission needs to provide incentives that will encourage scientists to embrace, rather than reject, this portentous but hazily defined future.

With that in-tray, the commissioner needs to be the kind of individual who genuinely believes that he or she can make a difference — and will rise to the challenge of doing so.

Ministers and commissioners tend to claim to have such ambitions when they begin. But their

default mode is usually that of a passenger, carried along by officials and events. Sadly, the departing commissioner, Máire Geoghegan-Quinn, falls squarely into this category.

It does not have to be that way. It is still possible for stout-hearted individuals such as Neelie Kroes, the current digital-agenda commissioner, to exert real influence.

The most recent research commissioner to leave a large footprint was, unfortunately, Édith Cresson, a former French prime minister, whose abuse of the position led to her 1999 resignation and subsequent conviction before the European Court of Justice in 2006. Every scientist subjected to the commission's hair-raising auditing process since then knows the true cost of Cresson's legacy.

Research and innovation is now the third-largest programme in the EU (after agriculture, and structural funds to aid poor regions). A real leader who knows the ropes politically, and has a clear agenda for European research from day one, could change the mood music for all of European research. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@gmail.com

IT IS STILL
POSSIBLE FOR
**STOUT-
HEARTED**
INDIVIDUALS TO
EXERT REAL
INFLUENCE.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/yisjnt

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

Fresh look at Galactic rim

A survey has provided the most detailed look yet at a mysterious ring of stars at the fringes of the Milky Way.

Using data from the Pan-STARRS1 telescope in Hawaii, Colin Slater and Eric Bell at the University of Michigan in Ann Arbor and their colleagues show that the Monoceros Ring appears as wispy stellar streams emerging from the Milky Way's outer disk.

There is debate over how the ring was created, with theories suggesting that it is either a part of the Galactic disk that was warped by the influence of nearby dwarf galaxies or the remnants of a dwarf galaxy that was unfurled in an encounter with the Milky Way. Neither scenario explains all the details seen in the survey, however, suggesting that the models need improving.

Astrophys. J. 791, 9 (2014)

NEUROSCIENCE

Brain scans predict TV hits

Brain activity measured in just a few individuals watching television programmes might predict whether large populations of viewers will find the shows interesting.

Jacek Dmochowski at Stanford University in California and his colleagues used functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) activity to follow brain activity in groups of up to 16 young adults watching a previously aired episode of drama programme *The Walking Dead*

(pictured) or advertisements broadcast during American football Super Bowl games.

The extent to which neural responses to the stimuli were shared between the small experimental groups correlated with the amount of social-media activity or positive audience ratings that the broadcasts had originally elicited from large audiences. Such neural reliability may be a useful tool in targeting



education or marketing activities to specific groups, the authors suggest.

Nature Commun. 5, 4567 (2014)

OCEAN SCIENCES

Breaking icebergs blast out noise

Iceberg disintegrations make the oceans noisier for months, and in the low frequencies that might affect marine mammals.

Researchers led by Haru Matsumoto at Oregon State University in Newport studied ocean hydrophone recordings from across the Southern Hemisphere. They found that

noise levels rose throughout the southern Pacific Ocean for 1.5 years after two huge icebergs disintegrated near Antarctica between 2007 and 2009. The signal was detected even north of the equator. *Geochem. Geophys. Geosys.* <http://doi.org/txf> (2014)

BIOENGINEERING

Rodents made see-through

The whole body of a rodent can be rendered transparent for imaging, without damaging cells and proteins.

Previous techniques for



CLIMATE SCIENCE

Rubbish is a burning problem

Open burning of rubbish contributes more than a trillion kilograms of carbon dioxide and other greenhouse gases to the atmosphere, but is often not included in national emissions estimates.

Christine Wiedinmyer, at the National Center for Atmospheric Research in Boulder, Colorado, and her colleagues estimated global waste-burning emissions on the basis of factors such as national population sizes, income and waste production and collection. They calculated that the CO₂ generated by open

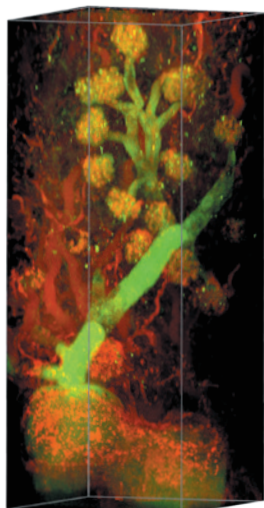
waste burning is equivalent to 5% of reported global anthropogenic emissions in 2010. In some countries, such as Mali and Sri Lanka, these emissions exceed those reported by the United Nations.

Emissions from burning rubbish are currently not accounted for in climate and air-quality models, so could explain some discrepancies between observed levels of pollutants and those estimated by models.

Environ. Sci. Technol. <http://doi.org/txh> (2014)

BEN HORTON/NATIONAL GEOGRAPHIC/CORBIS

AMC-TV/THE KOBAL COLLECTION



making tissue transparent have tended to work only on specific organs, such as the brain. A team led by Viviana Gradinaru at the California Institute of Technology in Pasadena tweaked an existing technique that stabilizes tissue and strips out light-blocking lipids with a cocktail of chemicals pumped through a dead rodent's circulatory system. Crucially, the process maintained the integrity of mouse neurons, kidney structures (**pictured**) and other tissue.

After one week for mice and two for rats, the brain and internal organs were clear and could be imaged under a microscope. The technique could allow researchers to see connectivity between the brain and other organs.

Cell <http://doi.org/tzz> (2014)

ZOOLOGY

Manta rays change colour

Csilla Ari of the University of South Florida in Tampa observed five manta rays at the Atlantis Aquarium in the Bahamas, and discovered white markings appearing and disappearing in the space of a few minutes on their backs, fins and heads.

The changes seemed to occur in response to feeding or interaction with other manta rays, and may represent a form of communication. Two of the animals were giant manta rays (*Manta birostris*; one

of two currently recognized species), and the other three may belong to a possible third species, which is similar but distinct from *M. birostris*.

Body coloration is used to identify species and individual rays, so the author says that understanding these colour changes is essential.

Biol. J. Linn. Soc.

<http://doi.org/txc> (2014)

ASTROPHYSICS

Best gauge of exoplanet size

Astronomers have made the most precise measurement so far of an exoplanet's size — for Kepler-93b, which orbits a star around 100 parsecs away.

Sarah Ballard at the University of Washington in Seattle and her colleagues estimated the planet's diameter at about 18,800 kilometres (1.48 times that of Earth), plus or minus 240 kilometres.

They used NASA's Kepler space telescope to monitor seismic activity inside the planet's host star. They also used the Spitzer Space Telescope to observe Kepler-93b as it transited the star, applying a technique that ensured that for each measurement, light from the star fell on the centre of the same pixel in Spitzer's camera.

This allows for precision measurements of exoplanets' radii and masses, and even of the structure of their parent stars, the authors write.

Astrophys. J. 790, 12 (2014)

ASTROPHYSICS

Novae join the γ -ray generators

Astronomers have identified a previously unknown source of cosmic γ -radiation.

High-energy γ -rays are released in extremely energetic events such as pulsars and supernovae. But they were thought to be unlikely products of classical novae: explosions that occur on the surfaces of compact, burnt-out stars called white

SOCIAL SELECTION

Furore over genome function

Just how much of our genome serves a purpose? A recent study has reignited this debate on social media. After comparing the genomes of 12 different mammals (including humans, mice and pandas), researchers at the University of Oxford, UK, concluded that only about 8.2% of the human genome is shaped by natural selection. The rest, they argue, is non-functional. Observers noted the large difference between this estimate and a previous claim by the ENCODE (Encyclopedia of DNA Elements) Project that 80% of the genome is biochemically active. Patrik D'haeseleer, a computational biologist at Lawrence Livermore National Laboratory, California, tweeted: "Only between 8% and 80% of human #genome is functional. Glad we've got that sorted out." At the heart of the issue are differing definitions of 'function'. Erick Loomis, an epigeneticist at Imperial College London, tweeted: "Maybe we should stop using 'functional' if we can't find a common definition."

PLoS Genet. 10, e1004525 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

NATURE.COM
For more on popular papers:
go.nature.com/boevfo

dwarfs as they collect material from their neighbours in the binary system. The only nova previously seen emitting such rays came from an unusual type of star system. Now Teddy Cheung at the Naval Research Laboratory in Washington DC and his colleagues have used NASA's Fermi Telescope to detect high-energy γ -rays coming from three classical novae.

The otherwise unremarkable properties of the three stars suggest that such emissions could be common. It is not yet clear how particles surrounding the stars might be accelerated enough to produce the energetic radiation.

Science 345, 554–558 (2014)

MICROBIOLOGY

A year with your microbes

Microbial communities in the gut and mouth have been followed every day for an entire year. Stool and saliva samples collected

from two men show that the communities remain fairly stable, but can be rapidly and broadly disrupted by events such as a bout of food poisoning or a holiday to a different continent.

Eric Alm, at the Massachusetts Institute of Technology in Cambridge, and his colleagues analysed these samples as well as health and lifestyle variables such as fitness, diet, exercise and mood recorded by the two volunteers.

One of the men developed food poisoning, which wiped out most of his gut bacteria; the microbes were eventually replaced with genetically similar species. And some lifestyle changes perturbed specific organisms — increasing dietary fibre, for instance, affected the abundance of 15% of the microbes in the gut.

Genome Biol. 15, R89 (2014)

NATURE.COM
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Study retracted

A landmark paper proposing a link between an influenza vaccine and narcolepsy was retracted on 31 July. The study, published last year (A. K. De la Herrán-Arita *et al. Sci. Transl. Med.* **5**, 216ra176; 2013), reported that some people with narcolepsy had immune cells that target a wakefulness-maintaining neurotransmitter. The cells also recognize some components of flu vaccines, it said, and the results explained why some children in Europe developed narcolepsy after receiving a vaccine for H1N1 swine flu. But the team discovered that it could not reproduce its own findings, and so retracted the paper. See go.nature.com/hbrrvi for more.

DOE opens access

The US Department of Energy (DOE) is making papers written by the researchers it funds freely available, it said on 4 August. A DOE online portal will link to peer-reviewed manuscripts and final-text journal papers within 12 months of their publication. Up to 30,000 studies are expected to be made available annually. The department is the first US federal agency to respond to orders for public access and data-sharing issued by the government 18 months ago. See go.nature.com/jp5owp for more.

Stellar survey

The European Space Agency's Gaia space telescope is ready to begin its five-year survey of about 1 billion stars in the Milky Way, the agency said on 29 July. Gaia will produce the most accurate three-dimensional map of the Galaxy yet. As it orbits the Sun, it will measure distances to stars by recording tiny shifts



ISAAC KASAMANI/AFP/GETTY

Anti-gay law overturned on technicality

Uganda's anti-gay law, which punishes some homosexual behaviour with life in prison, was nullified by a court in Kampala on 1 August. The move could help efforts to study HIV transmission and control the spread of the virus. Gay people in Uganda and other African nations are frequently unable to access

information on HIV, and those who become infected are often denied treatment (see *Nature* **509**, 274–275; 2014). The court ruled that the Anti-Homosexuality Act was invalid because too few members of parliament voted. It is not yet clear whether Uganda's government will appeal the ruling.

in their positions, and will observe the stars' movement through space. The telescope launched on 19 December but operations were delayed by light leaking into the detector, where it could have degraded observations of the faintest targets.

EVENTS

Hacker attack

Information technology (IT) systems at Canada's National Research Council have come under a cyberattack, the agency said on 29 July. The Canadian government blamed the intrusion on Chinese state-sponsored hackers. No details were provided about what data were accessed, but the council told scientists to expect disruptions. It

said it is overhauling its IT infrastructure and security, including integrating the system with the broader government network to help protect against future attacks. This work could take around one year to complete.

Ebola emergency

Officials are stepping up efforts to contain the West African Ebola outbreak, which had killed 887 people as of 1 August. The president of Sierra Leone declared a state of emergency on 30 July, allowing the police and military to quarantine infected homes and villages. The World Health Organization (WHO) also announced a US\$100-million plan to boost the number of emergency-response staff and scheduled

a meeting for 6 August to discuss the international implications of the epidemic.

FACILITIES

Space instruments

NASA's Mars 2020 rover mission will carry seven instruments to collect rocks for transport back to Earth. On 31 July NASA announced the winning instruments, which were chosen from 58 competitors. They include a zoomable camera, a machine to generate oxygen from carbon dioxide, and radar to explore geology up to half a kilometre deep. On 30 July, NASA also announced that the International Space Station will get two new instruments to observe how changes in Earth's climate and land use affect

HANS SAUTTER how forests and ecosystems function. See go.nature.com/zrbun7 for more.

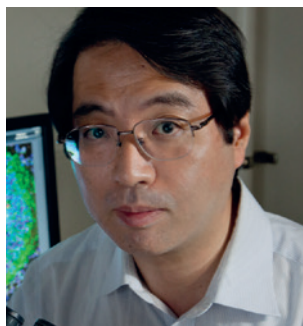
Neutrino detector

An international collaboration to build a neutrino detector in China was announced on 30 July in Beijing. The Jiangmen Underground Neutrino Observatory (JUNO) project, led by the Institute of High Energy Physics in Beijing, will bring together researchers from countries including France, Russia and the United States. The detector will study neutrinos coming from supernovae, Earth and nearby nuclear reactors. It will be the world's largest liquid scintillator detector, which captures luminescence when a neutrino interacts with atomic nuclei in the liquid, and aims to give the first measurement of the relative masses of the three known types of neutrino. The facility will be completed by 2020.

PEOPLE

Stem-cell suicide

One of Japan's top stem-cell researchers, Yoshiki Sasai (pictured), died on 5 August in an apparent suicide. The 52-year-old researcher, who worked at the RIKEN Center for Developmental Biology in Kobe, was best known for coaxing embryonic stem cells to differentiate into various



types of mature cells. In 2011, he stunned the world by mimicking an early stage in the development of the eye *in vitro* using embryonic stem cells. But over the past six months he became caught up in the controversy surrounding two *Nature* papers that claimed embryonic stem cells could be created through a method called stimulus-triggered acquisition of pluripotency (STAP). The papers were retracted on 2 July after evidence of misconduct was found. Sasai, a co-author, was cleared of direct involvement but was criticized for poor oversight of research. "The world scientific community has lost an irreplaceable scientist," said RIKEN president Ryoji Noyori. See go.nature.com/etrboi for more.

Director resigns

The director of the US National Institute of Neurological Disorders and

Stroke (NINDS) is stepping down, the institute announced on 31 July. Story Landis, who has headed NINDS since 2003, had prominent roles in programmes such as the BRAIN Initiative and the National Institutes of Health's programme to improve the reproducibility of science. Landis leaves at the end of September, when the deputy director of NINDS, neurologist Walter Koroshetz, will take over as acting director.

POLICY

Test clampdown

Hospitals and laboratories in the United States will soon no longer be able to design their own diagnostic tests without input from the US Food and Drug Administration (FDA). On 31 July, the FDA announced that it will regulate the development of diagnostic tests for various diseases, as well as genetic tests used to identify patients who may react to certain treatments. The regulations will be phased in over the next nine years and will prioritize tests for which an incorrect diagnosis could result in significant harm to a patient. See page 5 for more.

Lab safety

Laboratories working with hazardous chemicals must develop a culture of safety rather than just relying on

COMING UP

10–14 AUGUST

The American Chemical Society autumn meeting in San Francisco, California, has the theme 'chemistry and global stewardship'.
go.nature.com/c2to6u

10–15 AUGUST

The Ecological Society of America conference in Sacramento, California, includes how fire affects ecology as a key topic for discussion.
www.esa.org/am

compliance with regulations, says a report released on 31 July by the US National Research Council. The report was motivated by a series of recent high-profile accidents in university labs (see *Nature* 493, 9–10; 2013). It recommends better training, proactive analysis of hazards and rewards for researchers who take precautionary measures.

BUSINESS

Biotech job cuts

Biotechnology firm Amgen, of Thousand Oaks, California, announced on 29 July that it will cut up to 2,900 jobs — around 15% of its global workforce. The company said it could not yet specify how many of the lost jobs will be in research and development. But it did say that it will close facilities in Washington state and Colorado, which include research and manufacturing sites. The cuts will begin later this year and follow weak sales of the company's anaemia drug Aranesp (darbepoetin alfa). Amgen spent US\$979 million on research and development in the second quarter of this year, nearly 19% of its sales revenue.

➔ NATURE.COM

For daily news updates see:
www.nature.com/news

TREND WATCH

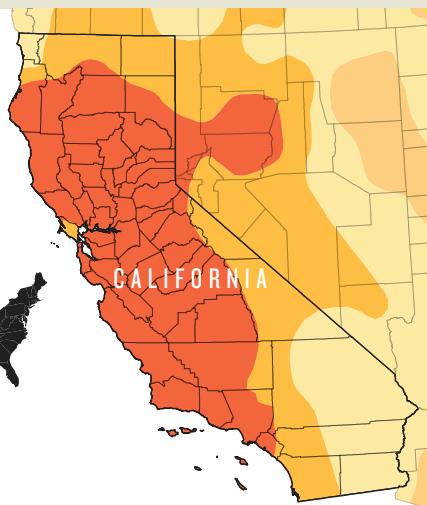
More than 20% of California shifted from extreme to exceptional drought — the most severe category — in the week up to 29 July. An update from the US Drought Monitor shows that exceptional drought now affects more than half of the state (see map), with more than 80% classified as under extreme drought or worse. California is short of more than a year's worth of reservoir water. "We wouldn't have expected it to be this dry," says a spokesman for the US Department of Agriculture.

CALIFORNIA DROUGHT

58% of the state is now under conditions of 'exceptional drought'.



Drought intensity in California
 ■ Severe
 ■ Extreme
 ■ Exceptional



NEWS IN FOCUS

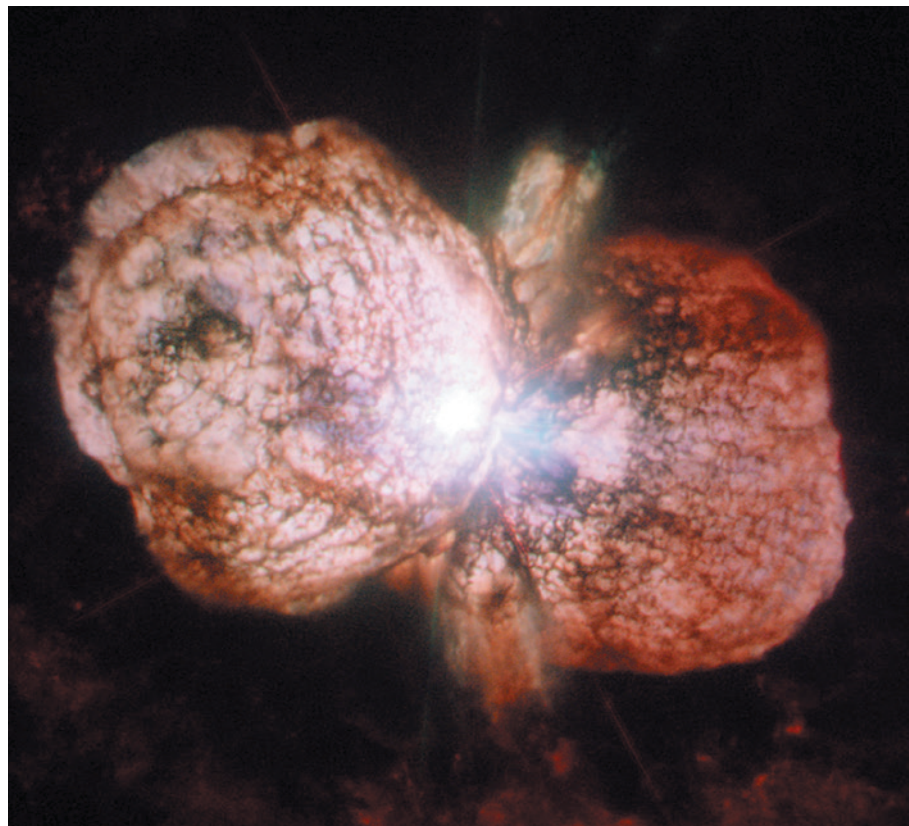
INFECTIOUS DISEASE Vaccine roll-out helps to stem tide of antimicrobial resistance **p.14**

PUBLISHING Open-access directory clamps down on predatory journals **p.17**

HEALTH CARE Heart disease chosen as subject of big-data clinical study **p.18**

CHEMISTRY The race to build the ultimate synthesis machine **p.20**

ESA/HUBBLE/NASA



The massive binary-star system η Carinae resembles the first stars that formed in the early Universe.

ASTRONOMY

Binary star to spill celestial secrets

Close approach and violent interaction of stars in η Carinae system will provide rare insight into stellar enigma.

BY ALEXANDRA WITZE

After centuries of perplexing scientists with its wildly erratic behaviour, a nearby star may give up some of its secrets in the next couple of weeks.

A binary system, η Carinae has two stars that swing past one another every 5.5 years. The

bigger star — some 90 times the mass of the Sun — is incredibly unstable, always seemingly on the verge of blowing up. When the smaller companion star makes its closest approach to the primary star, as is happening now, the interaction between the two triggers violent changes in the high-energy radiation pouring out of the system.

Astronomers are watching the show in the hope of learning what drives this enigmatic system. In the 1840s, η Carinae had a mysterious eruption; in recent decades, it has again brightened unexpectedly (see ‘Stellar show’). “The star is in an awfully deranged state, and no one knows why,” says Kris Davidson, an astronomer at the University of Minnesota in Minneapolis.

Some answers may come in the next few weeks. Theoretical work suggests that when η Carinae’s secondary star passes by, its fast stellar winds bore a huge hole into the outer layers of the primary star (T. I. Madura *et al. Mon. Not. R. Astron. Soc.* **436**, 3820–3855; 2013). Astronomers expect that if they are right about this, then a specific series of events will unfold this month — including a quick rise in the system’s X-ray production after a drop that began in mid-July.

Studying η Carinae has implications far beyond understanding one peculiar celestial system. Uncovering its secrets could help researchers to better understand the earliest stars that winked into existence. η Carinae is similar in mass to the first stars that formed in the Universe, billions of years ago. Most of today’s stars are much lighter, so η Carinae is a rare modern example of how such a massive star might operate — at the highly observable distance of 2,300 parsecs from Earth.

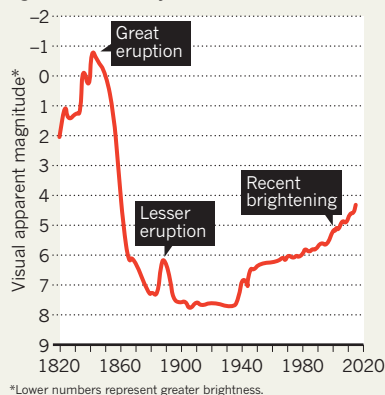
Across the Southern Hemisphere, professional and amateur astronomers are pointing their telescopes at the star, in the constellation Carina. “It’s the biggest effort ever,” says Theodore Gull, an astrophysicist at NASA’s Goddard Space Flight Center in Greenbelt, Maryland.

No one knows exactly when or how η Carinae’s companion will make its closest approach, but by mid-August it is likely to pass the primary at a distance equivalent to that between Mars and the Sun. Both stars in η Carinae emit powerful stellar winds, which at close range collide, producing a ‘bow shock’ like that seen in front of ships. The mutual tangle sets off a sequence of bizarre events.

The stars began brightening in the visible part of the electromagnetic spectrum in April, and then again in a sharper peak beginning in mid-June — probably as the companion star approached and began interacting with the primary star’s winds, says Eduardo Fernández Lajús, an astronomer at the National University of La Plata in Argentina. The system’s ►

STELLAR SHOW

Scientists have struggled to explain the erratic behaviour of the binary star η Carinae, which brightened unexpectedly in the 1840s and again more recently.



► X-ray production peaked in mid-July and has since plummeted to near zero — probably as the colliding winds, where the X-rays are born, have become entirely unstable and collapsed.

The Hubble Space Telescope and other instruments are also tracking dramatic changes in the chemical-element signatures found in η Carinae's light spectrum. The interaction between the two approaching stars can strip electrons from elements such as iron and helium, ionizing them more strongly than in normal celestial environments. "You have these bare helium nuclei — that's awfully hard to make in normal circumstances," says Gull. Watching this process over time helps to reveal how the stellar winds are interacting.

At the Pico dos Dias Observatory in southern Brazil, astronomer Augusto Damineli has been spending every night since 25 July trying to catch a glimpse of η Carinae through the winter clouds. On 29 July, his team finally caught a brief opening and managed to gather data showing that a helium spectral line is dropping in just the pattern that Damineli expected. "TOUCH-DOWN!" he wrote in an e-mail.

In 2009, when η Carinae had its most recent close encounter, the system's X-ray production plunged and then shot back up in half the time it did in 2003. That could be because the primary star's winds are slowing down, so it takes less time for the whole system to recover. If the wind speeds have continued to drop, X-ray emissions might shoot up even faster than last time.

Seeing such big differences from one close encounter to the next is "what everyone is waiting for," says Andrea Mehner, an astronomer at the European Southern Observatory in Santiago, Chile, who is monitoring η Carinae with Hubble. "We cannot make the star do something exciting if it doesn't want to." ■



A nurse prepares to immunize a young child with pneumococcal and rotavirus vaccines in Ghana.

VACCINES

Hidden bonus from vaccination

Immunization against pneumococcus in Africa also reduces levels of antibiotic resistance.

BY EWEN CALLAWAY

This summer, Eritrea, Côte d'Ivoire and Niger will join a growing list of countries where infants receive a vaccine to prevent pneumonia, meningitis and other deadly diseases caused by the pneumococcus bacterium (*Streptococcus pneumoniae*). Pneumonia is a leading killer of young children in low-income countries; vaccinations from 2010 to the end of this year are estimated to have averted 500,000 deaths, according to the GAVI Alliance in Geneva, an international

organization that facilitates vaccination.

Data from South Africa also point to another benefit of vaccination: stemming a rising tide of antibiotic resistance in the developing world. The country's introduction of a pneumococcal conjugate vaccine (PCV) in 2009 has not only reduced the overall incidence of invasive pneumococcal disease by about two-thirds in infants (the age group vaccinated) and in adults, but has also reduced penicillin-resistant infections in both groups.

This is the first time such benefits have been observed outside the developed world.

SOURCE: GAVI ALLIANCE

The data should spur public-health officials in low-income countries that have not yet adopted the vaccine to start using it, says Anne von Gottberg, a clinical microbiologist at the National Institute for Communicable Diseases in Johannesburg and leader of the study (see 'Protecting children'). Her group has reported the results at conferences but they have not yet been published.

The problem of antibiotic resistance is particularly stark in low-income countries, where over-prescription and poor regulation combine with a higher disease burden and poor sanitation to increase the use of antimicrobial drugs. A recent survey by the World Health Organization found rates of resistance in *Klebsiella pneumoniae* as high as 54%. Reduced susceptibility of *Streptococcus pneumoniae* to penicillin was found worldwide, and topped 50% in some reports.

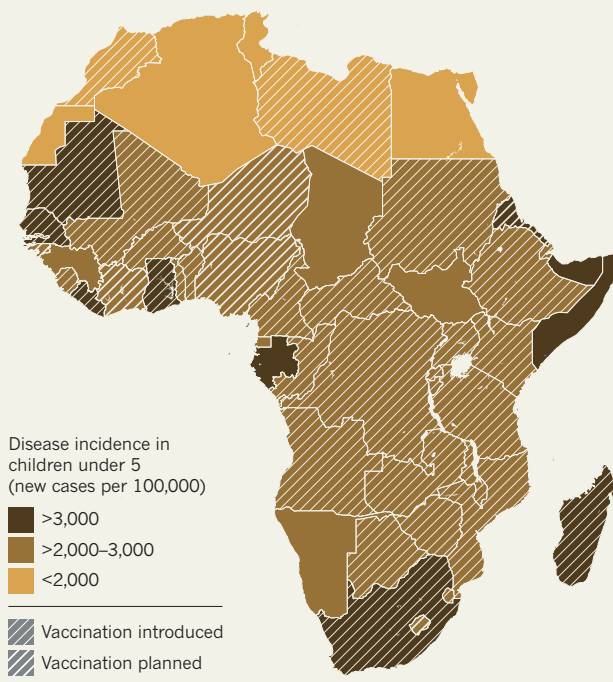
In North America, Europe and other well-off parts of the world, the introduction of pneumococcal vaccines in the early 2000s reduced cases of invasive pneumococcal disease by more than one-third in vaccinated children and in unvaccinated adults, who typically acquire infections from children. The vaccine also reduced the numbers of serious pneumococcal infections that were resistant to front-line antibiotics such as penicillin.

Between 1998 and 2008, a study in the United States found a 64% decrease in antibiotic-resistant pneumococci among children and a 45% decrease among adults over 65 (L. M. Hampton *et al. J. Infect. Dis.* **205**, 401–411; 2012). The different pneumococcal vaccines target a handful (7, 10 or 13, depending on the vaccine) of the more than 90 varieties (serotypes) of the pneumococcus, but those serotypes are among the most likely to develop antibiotic resistance. The result is a greater reduction in antibiotic-resistant strains in the population compared with sensitive strains.

Low-income countries began deploying pneumococcal vaccines around 2009, and

PROTECTING CHILDREN

Vaccination against the pneumococcus, a leading cause of death in young children, is gaining ground in Africa.



more than 40 of them are expected to administer the vaccine to infants by 2015. Many of these countries receive vaccines free or at a discount through financial support from the GAVI Alliance. "Getting those vaccines into low-income countries in a ten-year time span is an incredibly fast roll-out," says Katherine O'Brien, an epidemiologist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. She notes that most other vaccines have reached the world's poorest people much later after being introduced in developed countries.

South Africa, which is not eligible for GAVI support, funded its own pneumococcal vaccination programme in 2009 using the seven-strain PCV7, and von Gottberg's team tracked the effects. A different mix of pneumococcus serotypes circulates in South Africa from those in the Western countries for which the PCV7 vaccine was designed. And the types of people who contract pneumococcal disease also differ in South Africa, where HIV-infected adults, often mothers,

tend to contract the disease along with children and older people.

Despite these differences, von Gottberg's team noticed a steep reduction in rates of invasive pneumococcal disease after vaccination began. By 2012, cases caused by the serotypes included in the vaccine had plummeted in both children and in middle-aged adults (by 89% and 57%, respectively), and antibiotic-resistant infections with those serotypes were also down in the young (by 82%) and across all age groups. The results suggest that the vaccine's effects in Western countries translated to children and adults in a sub-Saharan African country. It is notable, von Gottberg says, that the incidence of drug-resistant cases fell more than the incidence of drug-sensitive ones.

O'Brien says that the data suggest that other Sub-Saharan African countries should experience similar benefits after introducing the vaccine. Ongoing surveillance of the effects of the PCV10 vaccine roll-out in 2010 in a district in

Kenya has also recorded steep reductions in severe pneumococcal infections (see go.nature.com/bxh6qp).

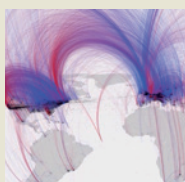
The reductions in antibiotic-resistant pneumococcal disease should lead to less use of antibiotics overall. A study in Finland found that the introduction of a different pneumococcal vaccine reduced antibiotic purchases by 8% (A. A. Palmu *et al. Lancet Infect. Dis.* **14**, 205–212; 2014). Moreover, if doctors are confident that front-line antibiotics will routinely cure serious pneumococcal diseases, "they won't feel that impetus to have to go with a big-gun antibiotic," says O'Brien. More prudent use of antibiotics should forestall the development of drug resistance.

Von Gottberg hopes that her team's data will encourage countries that haven't yet signed up for the vaccine to follow Eritrea, Niger and Côte d'Ivoire. "To have another weapon in our armament to reduce antibiotic resistance is a very good story," she says. "It will convince governments." ■



**MORE
ONLINE**

TOP STORY



Video map of births and deaths shows 2,600-year cultural history
go.nature.com/gwp54n

MORE ONLINE

- Siberian crater blamed on methane release go.nature.com/aoqyuv
- Five-year delay would doom ITER fusion-reactor experiment go.nature.com/ecdfpp
- Uncertainty over Cascadia quake risk go.nature.com/ima8nz

NATURE PODCAST



Straightening out nanotubes, chemistry's dream machine and an artist who fakes science nature.com/nature/podcast



Lars Bjørnshauge set up the Directory of Open Access Journals in 2003.

PUBLISHING

Open-access website gets tough

Leading directory tightens listing criteria to weed out rogue journals.

BY RICHARD VAN NOORDEN

When Lars Bjørnshauge founded a website to index open-access journals in 2003, just 300 titles made the list. But over the next decade, the open-access publishing market exploded, and Bjørnshauge's Directory of Open Access Journals (DOAJ) along with it. Today the DOAJ comprises almost 10,000 journals — and its main problem is not finding new publications to include, but keeping the dodgy operators out.

Now, following criticism of its quality-control checks, the website is asking all of the journals in its directory to reapply on the basis of stricter criteria. It hopes the move will weed out 'predatory journals': those that profess to publish research openly, often charging fees, but that are either outright scams or do not provide the services a scientist would expect, such as a minimal standard of peer review or permanent archiving. "We all know there has been a lot of fuss about questionable publishers," says Bjørnshauge.

The reapplication process will also create one of the largest 'whitelists' of acceptable open-access journals, helping the DOAJ to become a more useful tool for funders, librarians and researchers who want to look up information on a publication or import its metadata into their

catalogues. Those journals meeting the highest criteria — expected to be about 10–15% of the total — will also be given a 'seal' of best practice.

The DOAJ, which receives around 600,000 page views a month, according to Bjørnshauge, is already supposed to be filtered for quality. But a study by Walt Crawford, a retired library systems analyst in Livermore, California, last month (see go.nature.com/z524co) found that the DOAJ currently includes some 900 titles that are mentioned in a blacklist of 9,200

potential predatory journals compiled by librarian Jeffrey Beall at the University of Colorado Denver (see *Nature* **495**, 433–435; 2013). In addition, journalist John Bohannon last year proved that at least 73 journals in the DOAJ were suspect; in a sting operation, he sent them an obviously flawed paper which they then accepted for publication (J. Bohannon *Science* **342**, 60–65; 2013). The DOAJ removed the journals from its index.

The DOAJ had the idea of introducing stricter standards a few years ago, says Alma Swan, co-founder of the non-profit company IS4OA, which now operates the DOAJ (previously it was hosted by Lund University in Sweden). "We need to show which journals come up to a minimum standard of quality," she says.

Since May, would-be new members have had to fill in a tougher entry form containing more than 50 questions, which will now form the basis of the reapplication criteria. They include requests for information on a journal's digital archiving policy, its editorial board and its content licensing. "I suspect about 10% of journals on the list will not be able to pass the reapplication," says Bjørnshauge.

Paul Peters, the chief strategy officer at open-access publishers Hindawi, headquartered in Cairo, believes that the new criteria will be "incredibly important". "Scholarly researchers need a way to determine whether a given journal is adhering to best practice, and I believe that the DOAJ can provide a trusted and scalable mechanism for doing so," he says.

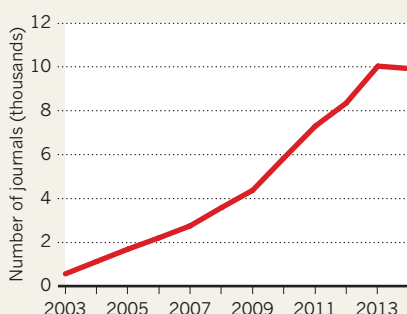
It is not clear whether the DOAJ's whitelist will become the pre-eminent index of trustworthy open-access journals. Beall says that the directory's credibility has already been hurt and that its new approach is "too little, too late". He is also not sure how the DOAJ will spot when a publisher is lying about its services. Moreover, Beall points out, many researchers and universities will instead judge a journal's quality by whether it is indexed in major citation databases, such as Elsevier's Scopus index, rather than looking at the DOAJ's list.

Bjørnshauge says that a small cohort of some 30 voluntary associate editors — mainly librarians and PhD students — will check the information submitted in reapplications with the publishers, and there will be a second layer of checks from managing editors. He also finds it "extremely questionable to run blacklists of open-access publishers", as Beall has done. (Crawford's study found that Beall's apparently voluminous list includes many journals that are empty, dormant or publish fewer than 20 articles each year, suggesting that the problem is not as bad as Beall says.)

But will any kind of whitelist help vulnerable researchers to avoid publishing in substandard journals? Beall doesn't think so. "There's no evidence that the whitelist approach has been helpful in encouraging researchers not to become victims of scams," he says. "Bad open-access publishers are still growing like crazy." ■

STUNTED GROWTH

The Directory of Open Access Journals grew rapidly — until it began culling low-quality publications last year.



US big-data health network launches aspirin study

PCORI clinical-research initiative will collect information on some 30 million people.

BY SARA REARDON

One of the largest big-data experiments in health care has set its first research target. The leaders of the Patient-Centered Outcomes Research Institute (PCORI) in Washington DC voted on 29 July to focus the institute's first clinical trial on the use of aspirin to prevent heart disease. The US\$10-million pilot study will be conducted through PCORnet, a network set up by PCORI to collect health-care data such as insurance claims, blood tests and medical histories for as many as 30 million people in the United States.

PCORnet is the latest — and largest — of a number of initiatives attempting to use patient data to improve health care. Private companies such as Kaiser Permanente in Oakland, California, and medical networks such as Partners HealthCare in Boston, Massachusetts, have similar aims. PCORI was established in 2010 with more than \$3 billion from the US government to fund research comparing the effectiveness of treatments, and has so far given out \$549 million in grants (see 'Network profile'). Its own project, PCORnet, comprises 29 smaller networks of US hospitals and patient groups. The aspirin trial is part of a pilot phase to work out questions such as how best to recruit study volunteers, standardize their records and build rapport with them.

In the second phase, scheduled to begin in September 2015, outside scientists will be able to mine PCORnet data for their research. "My measure of success will be the number of questions we get asked," says Christopher Forrest, a paediatrician at Children's Hospital of Philadelphia in Pennsylvania, who leads one of the networks participating in PCORnet.

The aspirin trial is scheduled to launch in early 2015. Participants will take daily doses of aspirin that fall within the range typically prescribed for heart disease, and be monitored to determine whether one dosage works better than the others. "While this study is a proof of concept, [aspirin dosage] is a very important question as well," says PCORI director Joe Selby. "We're proving that you can ask and answer questions that matter to patients, and this one has gone many years without an answer."

PCORI does not aim to test new potential treatments or uncover disease mechanisms — that falls within the purview of agencies such

DATA POINTS

Network profile

The Patient-Centered Outcomes Research Institute (PCORI) is setting up a database of US health records that will be used to compare the effectiveness of different medical treatments. The system, PCORnet, will connect multiple smaller networks, giving researchers access to records at a large number of institutions without creating a central data repository. Here are some key facts about PCORnet and its parent institution:

- Initial funding for PCORnet: **US\$93.5 million**
- Number of member networks: **29**
- Total number of patient records: up to **30 million**
- Budget for pilot study of aspirin dosage: **\$10 million**
- Number of projects funded by PCORI since 2012: **313**
- Amount of funding awarded by PCORI since 2012: **\$549 million**

as the US National Institutes of Health. "I see them as complementing, rather than replacing, randomized clinical trials," says William Hersh, a biomedical informatician at Oregon Health & Science University in Portland who is not involved in PCORI. One of the greatest challenges will be standardizing data from different networks to enable accurate comparison, he says. The many types of data — scans from medical imaging, vital-signs records and, eventually, genetic information — can be messy, and record-keeping systems vary between health-care institutions.

ADVICE AND CONSENT

Transparency in how patients' records are used will be essential for the programme to succeed, says Sam Smith, an activist at medConfidential, a UK-based organization that campaigns for patient privacy protection. "Generally if you tell people where their data are going, they're a lot happier than if you don't," he says. Transparency issues have plagued the UK National

Health Service's care.data, a medical-records database whose launch has been delayed amid concerns about data security and informed consent.

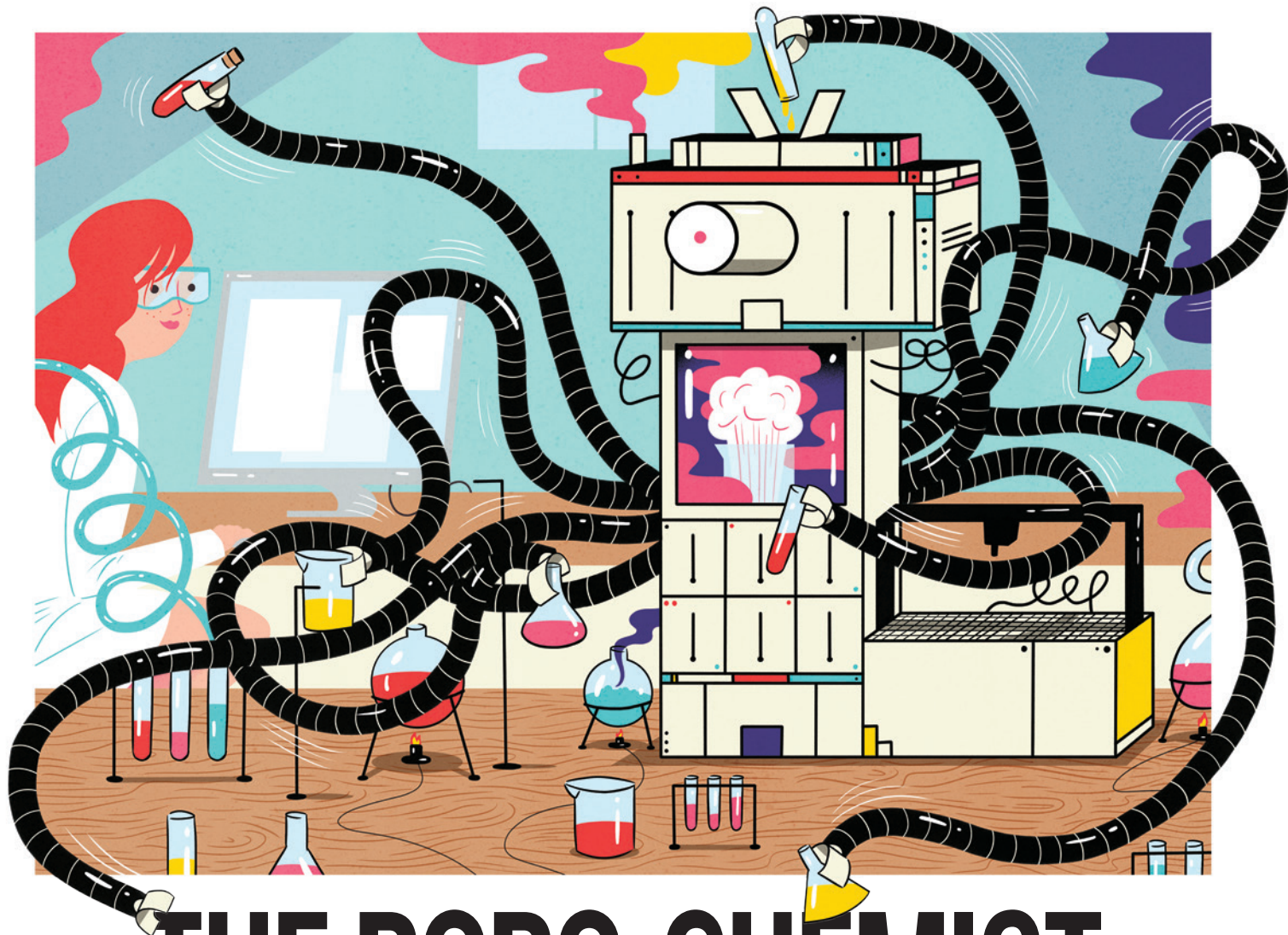
To avoid similar pitfalls, PCORI says that patient representatives will help to review its grant applications and work with investigators on trials. But simplifying scientific concepts for non-specialists can be hard, says Glenn Cohen, a bioethicist at Harvard University in Cambridge, Massachusetts. "It has a certain public spirit, but expecting patients to be involved raises challenges when these policy questions have a technical underlying element," he says. Selby says that PCORI plans to work out ways to deal with such issues over the next year.

To ensure privacy, PCORnet will not collect personal data. Patients' records will be kept by their health-care providers. When outside researchers use PCORnet for studies, the relevant data can be analysed within the network and the results sent to the researcher, or anonymized and provided in raw form.

The aspirin study will use standard procedures for obtaining informed consent, in which participants are told specifically about the research question their data will be used to address. But the process will be more complicated when outside researchers use PCORnet's data. People may tire quickly of signing forms every time a researcher wants to use their records for a study. "If very explicit consent is required, cost and practicality are probably going to make it a non-starter," Cohen says. The more transparency, patient involvement and data security that PCORI can provide, he says, the more ethically sound it would be to forgo standard informed-consent procedures. ■

CORRECTIONS

In the News story 'Biosafety controls come under fire' (*Nature* **511**, 515–516; 2014) the 'European Biosafety Organization' should have been the 'European Biosafety Association'. And the story 'Project drills deep into coming quake' (*Nature* **511**, 516–517; 2014) gave the old name for the Institute of Geological and Nuclear Sciences — it is now GNS Science.



THE ROBO-CHEMIST

The race is on to build a machine that can synthesize any organic compound. It could transform chemistry.

BY MARK PELOW

In faded photographs from the 1960s, organic-chemistry laboratories look like an alchemist's paradise. Bottles of reagents line the shelves; glassware blooms from racks of wooden pegs; and scientists stoop over the bench as they busily build molecules.

Fast-forward 50 years, and the scene has changed substantially. A lab in 2014 boasts a battery of fume cupboards and analytical instruments — and no one is smoking a pipe. But the essence of what researchers are doing is the same. Organic chemists typically plan their work on paper, sketching hexagons and carbon chains on page after page as they think through the sequence of reactions they will need to make a given molecule. Then they try to follow that sequence by hand — painstakingly mixing, filtering and distilling, stitching together molecules as if they were embroidering quilts.

ILLUSTRATION BY RYAN SNOOK

But a growing band of chemists is now trying to free the field from its artisanal roots by creating a device with the ability to fabricate any organic molecule automatically. “I would consider it entirely feasible to build a synthesis machine which could make any one of a billion defined small molecules on demand,” declares Richard Whitby, a chemist at the University of Southampton, UK.

True, even a menu of one billion compounds would encompass just an infinitesimal fraction of the estimated 10^{60} moderately sized carbon-based molecules that could possibly exist. But it would still be at least ten times the number of organic molecules that have ever been synthesized by humans. Such a device could thus offer an astonishing diversity of compounds for investigation by researchers developing drugs, agrochemicals or materials.

“A synthesis machine would be transformational,” says Tim Jamison, a chemist at the Massachusetts Institute of Technology (MIT) in Cambridge. “I can see challenges in every single area,” he adds, “but I don’t think it’s impossible”.

A British project called Dial-a-Molecule is laying the groundwork. Led by Whitby, the £700,000 (US\$1.2-million) project began in 2010 and currently runs until May 2015. So far, it has mostly focused on working out what components the machine would need, and building a collaboration of more than 450 researchers and 60 companies to help work on the idea. The hope, says Whitby, is that this launchpad will help team members to attract the long-term support they need to achieve the vision.

Even if these efforts fall short, say project members, early work towards a synthesis machine could still transform chemistry. It could deliver a host of reactions that work as continuous processes, rather than one step at a time; algorithms that can predict the best way to knit a molecule together; and important advances in how computers tap vast storehouses of data about the reactivity and other properties of chemicals. Perhaps most importantly, it could trigger a cultural sea change by encouraging chemists to record and share many more data about the reactions they run every day.

Some reckon it would take decades to develop an automated chemist as adept as a human — but a less capable, although still useful, device could be a lot closer. “With adequate funding, five years and we’re done,” says Bartosz Grzybowski, a chemist at Northwestern University in Evanston, Illinois, who has ambitious plans for a synthesis machine of his own.

ELECTRIC DREAMS

If chemists are to have any hope of building their dream device, they must pull together three key capabilities. First, the machine must be able to access a database of existing knowledge about how molecules can be built — which reactions create bonds between carbon atoms, for example, or whether using certain reagents to construct one part of a molecule risks damaging other parts. Second, it must be able to feed this knowledge into an algorithm that can map out synthetic steps, in much the same way that a master chess player plans a series of moves to win a game. And finally, it must be able to automatically carry out that sequence using real reagents inside a robotic reactor.

The technology for that last step has progressed the farthest. Many labs already own dedicated machines for churning out strands of DNA or polypeptides, and in the past decade, adaptable robot chemists have become increasingly important in commercial pharmaceutical research. But existing machines have limited capabilities: a DNA or protein sequence builder is typically able

to combine only a handful of molecular building blocks using fewer than half a dozen reactions. More versatile synthesis workstations are too expensive for most academic groups — costing from £30,000 to more than £500,000 — and still tend to produce molecules with a narrow range of chemical properties.

These workstations also do most of their reactions in the same batch-by-batch manner as humans. But some chemists are trying to develop continuous-flow synthesis, in which reactions occur as the chemicals move through the machine. This can improve speed and yields, and is a lot more amenable to automation.

Jamison, for example, is working on flow chemistry at the Novartis–MIT Center for Continuous Manufacturing in Cambridge, and he is part of a team that last year reported¹ the first end-to-end, completely continuous synthesis and formulation of a pharmaceutical: aliskiren hemifumarate, a treatment for high blood pressure. Jamison and his colleagues built a machine (now dismantled) that was more than 7 metres long, and about 2.5 metres high and deep. “It took four years of ‘everything that can go wrong, will go wrong,’” says Bernhardt Trout, head of the MIT centre and leader of the project. After a lot of trial and error, he says, the researchers got to the point at which they merely had to flip the switch and feed in fresh drums of solvent and raw

materials. The machine would hum like a large air-conditioning unit as stirrers whipped up chemicals, pumps whirled, filtration units dripped and squeezed, and a screw conveyor pushed solids through a 2-metre drying tube to be injection-moulded. Finally, after 14 operations and 47 hours, finished tablets dropped down a chute. Batch synthesis would have required 21 operations over 300 hours.

Jamison reckons that there is enormous potential for reactions to be adapted to continuous flow: “I think that it will be well over 50% eventually, maybe even 75%” of all reactions, he says. Progress is accelerating, he adds, because fixing a problem in one step — solids clogging a pipe, say — can offer immediate improvements to other processes.

A CHEMICAL BRAIN

Although automated machines are growing more versatile, teaching a computer to devise its own synthesis remains a massive problem, says Yuichi Tateno, an automation researcher at pharmaceutical company GlaxoSmithKline in Stevenage, UK, and a member of the Dial-a-Molecule collaboration. “The hardware has always been there, but the software and data have let it down,” he says.

Human chemists planning a synthesis tend to use a technique called retrosynthetic analysis. They draw the finished molecule and then pick it apart, erasing bonds that would be easy to form and leaving fragments of molecule that are stable or readily available. This allows them to identify the chemical jigsaw pieces they need as their raw materials, and to devise a strategy for connecting the pieces in the lab. If need be, they can seek inspiration from a commercial database such as SciFinder — an interface to the American Chemical Society’s Chemical Abstracts Service — or its main rival Reaxys, offered by publishing giant Elsevier. Entering a molecular structure or a reaction into these databases yields examples in the literature. But even with online help, says Tateno, humans often fail at synthesis. “With the amount of chemistry that’s out there, there’s nobody who can know it all.”

The hope is that a synthesis machine could one day do much better, says Whitby, not least because computers are so much faster at scanning through terabytes of chemical data to find a specific reaction. The bigger challenge, he adds, is that computers have a much harder time figuring

“A
**SYNTHESIS
MACHINE
COULD MAKE ANY OF
A BILLION
DEFINED SMALL
MOLECULES
ON DEMAND.**”

➔ **NATURE.COM**
For more about
synthesis machines:
go.nature.com/jrihfr

out whether that reaction will actually work in a synthesis, particularly if the target has never been made before.

That problem bedeviled Elias Corey, a chemist at Harvard University in Cambridge, Massachusetts, who formalized the rules of retrosynthesis in the 1960s. The following decade, Corey created software called LHASA (Logic and Heuristics Applied to Synthetic Analysis), which could use these rules to suggest sequences of steps towards a synthesis². But LHASA and its successors have never taken off, says Grzybowski: either the databases have included too few reactions and too many errors, or the algorithms have not properly assessed whether proposed reactions are compatible with all functional groups in the molecule. “If we could just make one chemical bond at a time, in isolation, chemistry would be trivial,” he says.

Grzybowski has spent the past decade building a system called Chematica to address those problems. He started by creating a searchable network of about 6 million organic compounds, connected by a similar number of reactions, drawn from one of the main databases behind Reaxys. His team then spent years cleaning up the data — identifying entries that lack crucial information about reagent compatibility or reaction conditions. Without that kind of clean-up, Chematica would be like a computer chef surveying a gigantic recipe book for dishes that use ice cream, stumbling on baked Alaska, and concluding that ice cream can withstand very high temperatures — missing the fact that cooking ice cream in an oven only works with an insulating shield of meringue. Chematica includes such crucial information, so its proposed syntheses of novel molecules — based on about 30,000 retrosynthetic rules — can be much more trustworthy.

The team also designed Chematica to take a holistic view of synthesis: it not only hunts for the best reaction to use at each step, but also considers the efficiency of every possible synthetic route as a whole. This means that a poor yield in one step can be counterbalanced by a succession of high-yielding reactions elsewhere in the sequence. “In 5 seconds we can screen 2 billion possible synthetic routes,” says Grzybowski.

STRONGER, FASTER, CHEAPER

When Grzybowski first unveiled the network behind Chematica in 2005 (ref. 3), “people said it was bullshit,” he laughs. But that changed in 2012, when he and his team published a trio of landmark papers^{4–6} showing Chematica in action. For example, the program discovered⁴ a slew of ‘one pot’ syntheses in which reagents could be thrown into a vessel one after the other, without all the troublesome separation and purification of products after each step. The group tested Chematica’s suggestions for making a range of quinolines — structures commonly found in drugs and dyes — and showed that many were more efficient than conventional approaches.

Chematica can also look up information about the cost of starting materials and estimate the labour involved in each reaction, allowing it to predict the cheapest route to a particular molecule. When Grzybowski’s lab tested 51 cut-price syntheses suggested by Chematica⁵, it collectively trimmed costs by more than 45%.

These demonstrations have impressed synthetic chemists, although few have had a chance to test Chematica. That is because Grzybowski is hoping to commercialize the system: he is negotiating with Elsevier to incorporate the program into Reaxys, and is working with the pharmaceutical industry to test Chematica’s synthesis suggestions for biologically active, naturally occurring molecules. Grzybowski is also bidding for a grant from the Polish government, worth up to 7 million złoty (US\$2.3 million), to use Chematica as the brain of a synthesis machine that can prove itself by automatically planning and executing syntheses

of at least three important drug molecules.

Others are doubtful that will happen — at least any time soon. For the foreseeable future, “there will always be a significant need for human intervention,” says Simon Tyler, commercial director of CatScl, a contract-research company in Cardiff, UK, that is involved in Dial-a-Molecule. “We won’t have RoboCops wandering around in the lab.”

And as long as programmes like Chematica rely on databases of published studies, says Whitby, they will struggle to design reliable synthetic routes to unknown compounds. To build a synthesis machine, “we need to be able to predict when a reaction is going to work — but more importantly we need to be able to predict when it’s going to fail”.

Unfortunately, those failures are rarely recorded in the literature. “We only publish the successes, a cleaned-up version of what happens in the lab,” says Whitby. “We also lose a lot of information: what really was the temperature, what was the stirring speed, how much solvent did you use?”

One solution is to record those successes and failures using electronic laboratory notebooks (ELNs), computer systems for logging raw experimental data that are widely used in industry but still rare in academia (see *Nature* **481**, 430–431; 2012). “A

lot of people ask, ‘Who reads all these data?’ The point is that machines use them — they can search the data,” explains Mat Todd, a chemist at the University of Sydney in Australia.

In principle, automated workstations and instruments could send information to an ELN, which would upload the details to an open-access database where they could help a synthesis machine to predict how reliable a reaction might be. “If we really did know the history of every chemical reaction that had ever been done, we’d have amazing predictive capabilities,” says Todd.

Dial-a-Molecule researchers have coordinated trials of ELNs in academic labs; started to devise a standardized, machine-readable format for ELN records; and developed software that can push those data into open databases such as ChemSpider. Others in the network have developed prototype software called PatentEye, which could pull in extra data by scraping and cataloguing chemical information from patents.

Many of those dreaming of a synthesis machine agree that widespread data harvesting will require a huge cultural shift. “That’s absolutely the biggest barrier,” says Todd. “In chemistry, we don’t have that culture of sharing, and I think it’s got to change.”

Money is also a significant hurdle. The expense of automated workstations means that few academics are familiar with them or their potential for capturing data. And with a large workforce of graduate students to draw on, academic labs often have little incentive to automate. Whitby is lobbying for a national centre that would host state-of-the-art automated synthesis equipment and software, to encourage their development and use. Until that materializes, he hopes that Dial-a-Molecule will inspire a new generation of chemists to embrace data sharing and automation.

Grzybowski, for one, is convinced that the synthesis machine can become a reality: “The only thing that can kill it is scepticism.” ■

Mark Peplow is a science journalist based in Cambridge, UK.

1. Mascia, S. *et al.* *Angew. Chem. Int. Edn* **52**, 12359–12363 (2013).
2. Corey, E. J., Howe, W. J. & Pensak, D. A. *J. Am. Chem. Soc.* **96**, 7724–7737 (1974).
3. Fialkowski, M., Bishop, K. J. M., Chubukov, V. A., Campbell, C. J. & Grzybowski, B. A. *Angew. Chem. Int. Edn* **44**, 7263–7269 (2005).
4. Gothard, C. M. *et al.* *Angew. Chem. Int. Edn* **51**, 7922–7927 (2012).
5. Kowalik, M. *et al.* *Angew. Chem. Int. Edn* **51**, 7928–7932 (2012).
6. Fuller, P. E., Gothard, C. M., Gothard, N. A., Weckiewicz, A. & Grzybowski, B. A. *Angew. Chem. Int. Edn* **51**, 7933–7937 (2012).

COMMENT

ART Albrecht Dürer's 16th century depiction of melancholy **p.26**



MENTAL HEALTH Back evidence-based therapies for treating depression **p.27**

EDUCATION University admissions policies should champion diversity **p.28**

RESEARCH ETHICS Developing rules for assessing pain in lab animals **p.28**

STEFAN CHRISTMAN/CORBIS



The aurora australis over the German Antarctic research base, Neumayer-Station III.

Six priorities for Antarctic science

Mahlon C. Kennicutt II, Steven L. Chown and colleagues outline the most pressing questions in southern polar research, and call for greater collaboration and environmental protection in the region.

Antarctica. The word conjures up images of mountains draped with glaciers, ferocious seas dotted with icebergs and iconic species found nowhere else. The continent includes about one-tenth of the planet's land surface, nearly 90% of Earth's ice and about 70% of its fresh water. Its encircling ocean supports Patagonian toothfish and krill fisheries, and is crucial for regulating climate and the uptake of carbon dioxide by sea water.

Antarctic scientists are unlocking the

secrets of Earth's climate, revealing lakes and mountains beneath the ice, exploring the deep sea and contemplating the origins of life and the Universe. Once seen as a desolate place frozen in time, Antarctica is now known to be experiencing relentless change. Local transformations such as the loss of ice, changes in ocean circulation and recovery of atmospheric ozone have global consequences — for climate, sea level, biodiversity and society.

In April 2014, the Scientific Committee on Antarctic Research (SCAR) convened

75 scientists and policy-makers from 22 countries to agree on the priorities for Antarctic research for the next two decades and beyond. This is the first time that the international Antarctic community has formulated a collective vision, through discussions, debate and voting. The SCAR Antarctic and Southern Ocean Science Horizon Scan narrowed a list of hundreds of scientific questions to the 80 most pressing ones (see Supplementary Information; go.nature.com/iilh5a). A full report will be published in August. ▶

► Here we summarize the overarching scientific themes, and outline steps that researchers and governments must take to make this vision a reality. Securing funding, as well as access to and protection for the region, will make greater international collaboration a necessity.

SIX SCIENTIFIC PRIORITIES

The questions identified fall broadly into six themes. To realize the full potential of Antarctic science we need to do the following.

Define the global reach of the Antarctic atmosphere and Southern Ocean. Changes in Antarctica's atmosphere alter the planet's energy budgets, temperature gradients, and air chemistry and circulation. Too little is known about the underlying processes. How do interactions between the atmosphere, ocean and ice control the rate of climate change? How does climate change at the pole influence tropical oceans and monsoons? How will the recovering ozone hole and rising greenhouse-gas concentrations affect regional and global atmospheric circulation and climate?

The Southern Ocean has important roles in the Earth system. It connects the world's oceans to form a global system of currents that transfers heat and CO₂ from the atmosphere to the deep ocean. Nutrients carried north support the base of the ocean's food web. The ocean is becoming more acidic as CO₂ dissolves in sea water, and cold southern waters will be the first to exhibit impacts. How will climate change alter the ocean's ability to absorb heat and CO₂ and to support ocean productivity? Will changes in the Southern Ocean result in feedbacks that accelerate or slow the pace of climate change? Why have the deepest waters of the Southern Ocean become warmer and fresher in the past four decades?

Sea ice reflects and filters sunlight. It modulates how heat, momentum and gases exchange between the ocean and atmosphere. Sea-ice formation and melt dictate the salt content of surface waters, affecting their density and freezing point. What factors control Antarctic sea-ice seasonality, distribution and volume? We need to know.

Understand how, where and why ice sheets lose mass. The Antarctic ice sheet contains about 26.5 million cubic kilometres of ice, enough to raise global sea levels by 60 metres if it returned to the ocean. Having been stable for several thousand years, the Antarctic ice sheet is now losing ice at an accelerating pace^{1,2}. What controls this rate and the effect on sea level? Are there thresholds in atmospheric CO₂ concentrations beyond which ice sheets collapse and the seas rise dramatically? How do effects at the base of the ice sheet influence its flow, form and

response to warming? Water bodies beneath the thick ice sheet have barely been sampled, and their effect on ice flow is unknown.

Reveal Antarctica's history. Glimpses of the past from rock records collected around the continent's margins suggest that Antarctica might look markedly different in a warmer world. But rocks from the heart of the con-

“Maximizing scientific return while minimizing the human footprint should be the goal.”

continent and the surrounding oceans have been only sparsely probed. Responses of the crust to, and the effects of volcanism and heat from Earth's interior on, overlying ice are largely undescribed. We know little about the structure of the Antarctic crust and mantle and how it influenced the creation and break-up of super-continents. Ancient landscapes beneath ice reveal the history of interactions between ice and the solid Earth. Geological signatures of past relative sea level will show when and where planetary ice has been gained or lost. We need more ice, rock and sediment records to know whether past climate states are fated to be repeated.

Learn how Antarctic life evolved and survived. Antarctic ecosystems were long thought of as young, simple, species-poor and isolated. In the past decade a different picture has emerged. Some taxa, such as marine worms (polychaetes) and crustaceans (isopods and amphipods) are highly diverse, and connections between species on the continent, neighbouring islands and the deep sea are greater than thought. Molecular studies reveal that nematodes, mites, midges and freshwater crustaceans survived past glaciations.

To forecast responses to environmental change we need to learn how past events have driven diversifications and extinctions. What are the genomic, molecular and cellular bases of adaptation? How do rates of evolution in the Antarctic compare with elsewhere? Are there irreversible environmental thresholds? And which species respond first?

Observe space and the Universe. The dry, cold and stable Antarctic atmosphere creates some of the best conditions on Earth for observing space. Lakes beneath Antarctic glaciers mimic conditions on Jupiter and Saturn's icy moons, and meteorites collected on the continent reveal how the Solar System formed and inform astrobiology.

We have limited understanding of high-energy particles from solar flares that are funnelled to the poles along the Earth's magnetic field lines. What is the risk of solar events disrupting global communications and power systems? Can we prepare for them and are they predictable?

Recognize and mitigate human influences. Forecasts of human activities and their impacts on the region are required for effective Antarctic governance and regulation. Natural and human impacts must be disentangled. How effective are current regulations in controlling access? How do global policies affect people's motivations to visit the region? How will humans and pathogens affect and adapt to Antarctic environments? What is the current and potential value of Antarctic ecosystem services and how can they be preserved?

CHALLENGING ENVIRONMENT

Answering these many questions will require sustained and stable funding; access to all of Antarctica throughout the year; application of emerging technologies; strengthened protection of the region; growth in international cooperation; and improved communication among all interested parties.

Antarctic programmes are sensitive to budget uncertainties and disruptions. In the past year, US projects were deferred, delayed or reduced in scale because of the US government shutdown in October 2013. Other national programmes suffered budget cuts stemming from the economic slowdown. High fuel prices and diversions for a major search and rescue mission hindered some. Decades-long projects are difficult to sustain given short grant cycles.

Postponed projects and lost field seasons leave gaps — a missing year of data for an ice-sheet study or biodiversity monitoring is irreplaceable. Faced with such uncertainties and hurdles, and with laboratories and students to support, some Antarctic researchers choose to leave the field. This also jeopardizes the recruitment and retention of the next generation of researchers.

Access to locations needed for science is limiting. Much of the continent and the Southern Ocean remain unexplored, and most scientists visit for a only few months each year. Researchers will need to develop autonomous vehicles and observatories that can reach remote locations such as beneath ice shelves, the deep sea and under ice sheets. Miniaturized sensors deployable on floats, animals and ice tethers must be able to acquire or transmit data for months or years.

A wider range of satellite-borne sensors is needed to continuously observe the entire region. Expanded aircraft-based geophysical surveys are needed to access the continental interior and ice margins. Advanced biogeochemical and biological sensors will be crucial for establishing regional patterns. Databases and repositories that can handle vast quantities of genomic and biodiversity information will be essential.

Future data sets will require high-speed and high-volume communications over great distances. Reliable sources of energy to power



Emperor penguins dive under a breathing hole in the Antarctic sea ice, which provides a platform for marine life.

remote observatories and better ways to store and uplink data will be needed. Improved computer models are essential for portraying the highly interconnected Antarctic and Earth system if we are to improve forecasts.

Antarctica's environmental-protection measures must be strengthened^{3,4}. More scientists will need to visit, and tourist numbers have almost tripled in the past decade to more than 34,000 a year plus support personnel. This growth increases the risk of introducing non-indigenous species and the likelihood of fuel spills that we are ill-equipped to respond to effectively^{3,5}.

The Antarctic Treaty System, which is responsible for governance of the region, is being tested by mounting environmental pressures and economic interests^{3,6}. The establishment of marine protected areas, international regulation of tourism, assessing financial penalties for environmental damage and regulating bioprospecting have proved difficult to resolve. An integrated strategy for Antarctic environmental management is essential⁴.

Antarctica is seen as a place to assert national interests⁶. In the past decade, countries including Belgium, China, the Czech Republic, India and South Korea have established new stations; Germany, the United Kingdom, the United States and others have replaced ageing ones; and Japan, South Korea and South Africa have built or replaced ice-capable ships.

Yet scientists from many other nations lack access to Antarctica. Twenty-nine countries participate in decision-making and another twenty-one have agreed to abide by the Antarctic Treaty. Although this

represents about two-thirds of the world population, it comprises less than one-sixth of the 193 member states of the United Nations — countries in Africa and the Middle East are notably under-represented.

WORK TOGETHER

Maximizing scientific return while minimizing the human footprint should be the goal. Coordinated international efforts that engage diverse stakeholders will be crucial.

It is time for nations involved in southern polar research to embrace a renewed spirit of cooperation as espoused by the founders of the Antarctic Treaty — in actions not just words. Wider international partnerships, more coordination of science and infrastructure funding and expanded knowledge-sharing are essential.

As an interdisciplinary scientific body, but not a funder of research, SCAR should assist with and encourage coordination and planning of joint projects, sharing of data and dissemination of knowledge to policymakers and the public. SCAR should repeat the Horizon Scan exercise every four to six years and provide the outcomes to emerging integrated science, conservation and policy efforts^{3,4} (see www.environments.aq).

We urge the Antarctic Treaty and its Committee for Environmental Protection to expand use of scientific evidence in its decision-making and to apply state-of-the-art conservation measures judged on measurable outcomes⁷.

Communicating the global importance of Antarctica to the public is a priority⁸. Narratives must better explain how the region

affects and is influenced by our daily lives. Antarctic success stories, such as signs of ozone recovery, engender confidence in the power of changes in behaviour.

Antarctic science is globally important. The southern polar community must act together if it is to address some of the most pressing issues facing society. ■

Mahlon C. Kennicutt II is professor emeritus of oceanography at Texas A&M University, College Station, Texas, USA, and former president of the Scientific Committee on Antarctic Research. **Steven L. Chown** is professor of biological sciences at Monash University, Victoria, Australia.

John J. Cassano, Daniela Liggett, Rob Massom, Lloyd S. Peck, Steve R. Rintoul, John W. V. Storey, David G. Vaughan, Terry J. Wilson, William J. Sutherland. e-mail: mckennicutt@gmail.com

1. Rignot, E., Mouginot, J., Morlighem, M., Seroussi, H. & Scheuchl, B. *Geophys. Res. Lett.* **41**, 3502–3509 (2014).
2. Joughin, I., Smith, B. E. & Medley, B. *Science* **344**, 735–738 (2014).
3. Chown, S. L. et al. *Science* **337**, 158–159 (2012).
4. Tin, T., Liggett, D., Maher, P. T. & Lamers, M. (eds.) *Antarctic Futures: Human Engagement with the Antarctic Environment* (Springer, 2014).
5. Chown, S. L. et al. *Proc. Natl Acad. Sci. USA* **109**, 4938–4943 (2012).
6. Brady, A.-M. (ed.) *The Emerging Politics of Antarctica* (Routledge, 2014).
7. Shaw, J. D., Terauds, A., Riddle, M. J., Possingham, H. P. & Chown, S. L. *PLoS Biol.* **12**, e1001888 (2014).
8. Rapley, C. et al. *Time For Change? Climate Science Reconsidered* (UCL, 2014).

On behalf of attendees at the 1st SCAR Antarctic and Southern Ocean Science Horizon Scan Retreat, 20–23 April 2014, Queenstown, New Zealand. See go.nature.com/iilhsa for a full list of co-signatories.

are joined in a web of hidden forces and symbols. The humour melancholy, for instance, is governed by the planet Saturn, whence the word 'saturnine'.

So there would have been nothing obscure about this picture for its intended audience of intellectual connoisseurs. Dürer mastered and exploited the new technologies of printmaking, so he was able to distribute his works widely, and he indicated in his diaries that he sold many and gave others as gifts to friends and humanist scholars such as Erasmus of Rotterdam. Ferdinand Columbus, son of Christopher, collected more than 3,000 prints, 390 of them by Dürer and his workshop.

Although the alchemical imagery of *Melencolia I* was part of this engraving's 'occult parcel', it would be wrong to imagine that alchemy was, to Dürer and his contemporaries, purely an esoteric art. As Lawrence Principe has argued in *The Secrets of Alchemy*

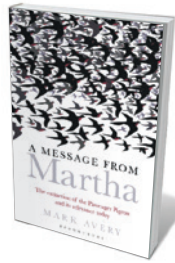
"There would have been nothing obscure about this picture for its intended audience of intellectual connoisseurs."

(University of Chicago Press, 2012), this precursor to chemistry was not only, or even primarily, about furtive and futile attempts to make gold from base metals. It was also a practical craft, not least in providing artists with their pigments. For this reason, artists commonly knew something of its techniques; Lucas Cranach the Elder, a friend of Dürer's, was a pharmacist on the side, which may explain why he was almost unique in northern Europe in using the rare and highly poisonous yellow pigment orpiment, an arsenic sulphide. The extent of Dürer's chemical knowledge is not known, but he was one of the first artists to use acids for etching metal, a technique developed only at the start of the sixteenth century. The process required specialist knowledge: it typically used nitric acid, made from salt-petre, alum and ferrous sulphate, mixed with 'Dutch mordant' composed of dilute hydrochloric acid and potassium chlorate.

Humility should perhaps compel us to concur with art historian Keith Moxey that "the significance of *Melencolia I* is ultimately and necessarily beyond our capacity to define" — we are too removed from it now for its themes to resonate. But what surely endures in this image is a reminder that for the Renaissance artist there was continuity between theories about the world, matter and human nature, the practical skills of the artisan, and the business of making art. ■

Philip Ball is a writer based in London. His latest book is *Invisible*.
e-mail: p.ball@btinternet.com

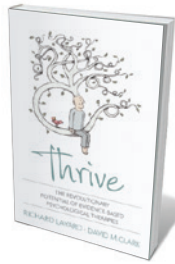
Books in brief



A Message from Martha: The Extinction of the Passenger Pigeon and its Relevance Today

Mark Avery BLOOMSBURY (2014)

A century ago, Martha — the last of the passenger pigeons (*Ectopistes migratorius*) — died in a zoo in Cincinnati, Ohio. Less than 100 years before, the species had been Earth's most common bird, darkening the sky and pulling down trees with the sheer size of its colonies. Conservationist Mark Avery's chronicle is based on science, historical accounts, a 6,000-kilometre road trip to key US sites, and numerous interviews. It offers a considered perspective on the habitat loss and unsustainable harvesting that led to the pigeon's demise.



Thrive: The Revolutionary Potential of Evidence-Based Psychological Therapies

Richard Layard and David M. Clark ALLEN LANE (2014)

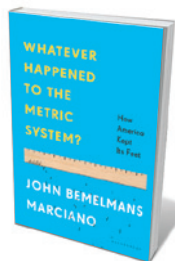
More than 350 million people worldwide have depression, estimates the World Health Organization. Yet mental illness is a policy blind spot and access to treatment is poor — a "shocking form of discrimination", say psychologist David Clark and economist Richard Layard. Drivers of the UK Improving Access to Psychological Therapies initiative, they make the case for tackling the burden now, and draw up a road map for evidence-based therapies recommended by the UK National Institute for Health and Care Excellence.



Extracted: How the Quest for Mineral Wealth Is Plundering the Planet

Ugo Bardi CHELSEA GREEN (2014)

Our dependence on fossil fuels and minerals is growing ever more costly, as extraction rapidly creams off the "easy" deposits. So argues physical chemist Ugo Bardi in this in-depth study of the issue, based on a report to the Club of Rome, a global think tank. Bardi examines depletion models, pollution and climate change, the viability of mining oceans or asteroids, and options such as waste recycling. If rampant extraction persists, he notes, we will not need spaceships to find a new world: we'll be standing on it, and it will not be pretty.



Whatever Happened to the Metric System?: How America Kept Its Feet

John Bemelmans Marciano BLOOMSBURY USA (2014)

Miles and pounds are here to stay — in the United States, at least. Its measurement system survived an abortive federal metric initiative in the 1970s and 1980s and, writer John Bemelmans Marciano reveals in this digressive history, a much earlier attempt. In 1790 Thomas Jefferson hoped to follow up his decimal currency with a decimal system of weights and measures. Instead, it was US ally post-revolutionary France, burdened with a hideously complicated system of weights and measures, that authored the metric system.



The Amazing World of Flyingfish

Steve N. G. Howell PRINCETON UNIVERSITY PRESS (2014)

The more than 60 species of the family Exocoetidae literally sail the seas — or, more accurately, just above them. The streamlined bodies of flying fish lend them underwater speed that they can convert into lift; winglike pectoral fins allow them to glide as far as 180 metres to escape predators. They can even use updraughts of air. Ornithologist Steve Howell's engrossing natural history is embellished with 90 superb colour photographs of the ornate goldwing and other beauties among these "ocean butterflies". **Barbara Kiser**

Correspondence

Private collections hold back science

Christian Foth and colleagues describe the spectacular eleventh specimen of the earliest bird, *Archaeopteryx*, including plumage features previously unknown for this pivotal taxon (*Nature* **511**, 79–82; 2014). We are concerned, however, about access to this important specimen after its long-term loan to a public museum expires, when it will be returned to a private collection.

This *Archaeopteryx* has been registered under the 'Act to Prevent the Exodus of German Cultural Property' (see go.nature.com/xyk5lz), which requires its whereabouts to be recorded and prevents loss of German heritage. However, public access to such specimens remains at the owner's discretion.

Many journals refuse to publish characteristics of specimens held in private collections because the observations cannot be independently verified, which is counterproductive for the scientific community. Instead, journals should insist on a guarantee of future access as a condition of publication.
Paul M. Barrett, Martin C. Munt *Natural History Museum, London, UK.*
p.barrett@nhm.ac.uk

Smart tools boost mental-health care

Portable electronic devices such as smart phones and virtual-reality headsets provide clinicians with a powerful tool for improving evidence-based psychological treatments (see *Nature* **511**, 287–289; 2014). For example, they offer patients more-frequent access to therapy and are likely to boost compliance.

Internet-based cognitive behavioural therapy, for instance, is effective for conditions such as depression

and anxiety. Immersive virtual reality helps for some phobias, including fear of flying and fear of heights, and holds promise for conditions such as eating disorders and post-traumatic stress disorder (see G. Riva *CyberPsychol. Behav.* **8**, 220–230; 2005).

Smartphones, in conjunction with wearable sensors, convey information about patients' activity, location and even physiological responses, providing insight into how well they function in everyday life and guiding decisions on mental-health interventions.

Virtual reality could also be a useful tool for researchers investigating the neural basis of functional and dysfunctional psychological processes (see C. J. Bohil *et al. Nature Rev. Neurosci.* **12**, 752–762; 2011).
Andrea Gaggioli, Giuseppe Riva *Catholic University; and Istituto Auxologico, Milan, Italy.*
andrea.gaggioli@unicatt.it

Rules for assessing pain in lab animals

New rules in the European Union directive for protecting laboratory animals call for a formal assessment of any pain that these animals might experience during scientific procedures (see go.nature.com/porg2x).

An upcoming workshop, organized by Germany's Federal Institute for Risk Assessment, will evaluate national guidelines on pain assessment drawn up in a similar workshop last year, with particular reference to 'harmful phenotypes' — animal lines that are likely to experience pain as a result of genetic alteration (see go.nature.com/8dhjcz).

Maintenance of such genetically altered animals, which carry the defining mutation at a single genetic locus, is currently subject to project authorization. This consent ensures that pain is monitored closely in these animals by veterinarians, animal-welfare officers, scientists

and laboratory staff who are responsible for animal care. Breeding more animals for the purpose of assessing pain severity is not permitted, and neither is subjecting them to further pain — as inflicted, for example, by routine blood testing.

Participants in the next workshop, to be held in October, will collate experiences resulting from implementation of the 2013 guidelines and will develop a database for pain-assessment results from genetically altered animals.

Barbara Grune, Andreas Hensel *Federal Institute for Risk Assessment (BfR), Berlin, Germany.*
Gilbert Schönfelder *BfR; and Charité-Universitätsmedizin Berlin, Germany.*
gilbert.schoenfelder@bfr.bund.de

Early life is key to disease risk

Lawrence O. Gostin calls for action to curb non-communicable diseases (*Nature* **511**, 147–149; 2014). Interventions in early life can also make a big difference.

There is now overwhelming evidence that factors in a child's early environment, during periods of developmental plasticity — including *in utero* — are major risk determinants for non-communicable disease in later life (see, for example, R. Barouki *et al. Environ. Health* **11**, 42; 2012).

We should heed recommendations, based on findings in the field known as developmental origins of health and disease, that preventive strategies during fetal development and early childhood will ultimately be more effective in reducing the long-term burden of disease. (see, for example, S. C. Davies *et al. Lancet* **382**, 1383–1384; 2013; and M. W. Gillman and D. S. Ludwig *N. Engl. J. Med.* **369**, 2173–2175; 2013).

Jeffrey M. Craig Murdoch

Childrens Research Institute, The Royal Children's Hospital, Parkville, Australia.
jeff.craig@mcri.edu.au
Susan Prescott *School of Paediatrics and Child Health, University of Western Australia, Perth, Australia.*

More inclusivity in PhD admissions

More PhD programmes in science and engineering should adopt the mentoring model developed by Casey Miller and Keivan Stassun to boost numbers of female and minorities students (see *Nature* **510**, 303–304; 2014). They should also use more-inclusive admissions criteria.

Efforts to democratize and diversify US higher education are placing a growing emphasis on applicants' social and economic origins (see, for example, W. G. Bowen *et al. Equity and Excellence in American Higher Education* Univ. Virginia Press, 2005). I suggest that admissions measures be expanded to include applicants who are first-generation college students, for example, or who were raised in poverty or in deprived single-parent families.

This broadened approach would fit with Miller and Stassun's recommendation that admissions practices be augmented with "proven markers of achievement, such as grit and diligence".

Kenneth Oldfield *University of Illinois at Springfield, Illinois, USA.*
koldf1@uis.edu

CORRECTION

In the Outlook article 'Funding by numbers' (*Nature* **511**, S52–S53; 2014), the vertical axis of the graphic 'Scholarly spending' was wrongly labelled. It should have read 'Percentage of GDP spent on research and development'.

Bias towards large genes in autism

ARISING FROM I. F. King *et al.* *Nature* **501**, 58–62 (2013); doi:10.1038/nature12504

In an important recent paper, King *et al.*¹ reported that inhibition of TOP1 and other topoisomerases reduces the expression of extremely long genes. They also showed that the list of large genes affected by TOP1 inhibition is enriched with candidate genes for autism spectrum disorders (ASD); however, the list of candidate genes that was used contains many genes with limited evidence for association with ASD². Here we demonstrate that the size of the genes among ASD candidate genes is biased towards extremely large genes only for genes identified to be disrupted by copy number variations (CNVs). Thus, our analysis suggests that the association between large genes and ASD is mainly driven by the method that implicated the genes in ASD. There is a Reply to this Brief Communication Arising by Zylka, M. J. *et al.* *Nature* **512**, <http://dx.doi.org/10.1038/nature13584> (2014).

The literature on ASD mentions many candidate genes, yet convincing evidence was yielded only for few of them. This is reflected in the SFARI Gene database used by King *et al.*¹. This database currently contains 588 genes, for which the evidence for association with ASD varies considerably. To address this concern, a gene scoring module was developed in SFARI Gene 2.0 to estimate the evidence level for individual genes². For example, Table 1 of King *et al.*¹ lists 49 ASD candidate genes that were affected by Top1 inhibitors, but only three of these genes are considered strong candidates, four genes have suggestive evidence, and an additional two genes are involved in syndromes associated with ASD. To retest the association between gene length and ASD, we first selected genes in the SFARI database that had a score of at least suggestive evidence. The gene with the strongest evidence, *CHD8*, is not particularly large (~50 kilobases), but the list also contains some of the largest genes in the genome like *AUTS2* and *NRXN1* (~1,000 kb). The SFARI Gene database is based on studies that focused on specific genes and on genome-wide studies, which are considered to be unbiased. However, as we show below, even in genome-wide studies there are biases influenced by the type of study.

The genome-wide search for ASD risk genes has been performed mainly by searching for rare and *de novo* variants, using two main methods: microarrays to identify CNVs, and exome sequencing to identify single nucleotide variations (SNVs) predicted to alter the protein sequence. Under the assumption of uniform mutation rate, the probability for a coding SNV is not strongly related to the total size of the gene, which is mainly determined by the length of the introns and untranslated regions. However, this is not true for CNVs. Most CNVs identified in ASD are large and contain multiple genes^{3,4}. Therefore, it is hard to associate any particular gene with ASD. In contrast, when a gene is extremely large there is a higher probability for it to become an ASD risk gene because it was the only gene affected by the CNV. Accordingly, we hypothesized that large genes will be associated with ASD mainly if implicated by CNVs.

Following the above hypothesis, we divided the ASD genes from the SFARI Gene database into two groups on the basis of the mutation type that implicated the gene in ASD: genes affected by CNVs or by SNVs. We plotted the distribution of gene sizes separately for each group. As can be seen in Fig. 1a, the distributions were notably different. “Genes with CNVs” showed bimodal distribution of short and large genes, whereas “genes with SNVs” were relatively short.

To further study the association between gene size and mutation types, we focused on six studies that searched for *de novo* CNVs and SNVs, mostly in the same cohort (the Simons Simplex Collection)^{3–8}. We compared the distribution of gene lengths between all coding genes in the genome, brain-specific genes, and genes with *de novo* SNVs or CNVs, in both ASD probands and unaffected siblings (Fig. 1b). Consistent with

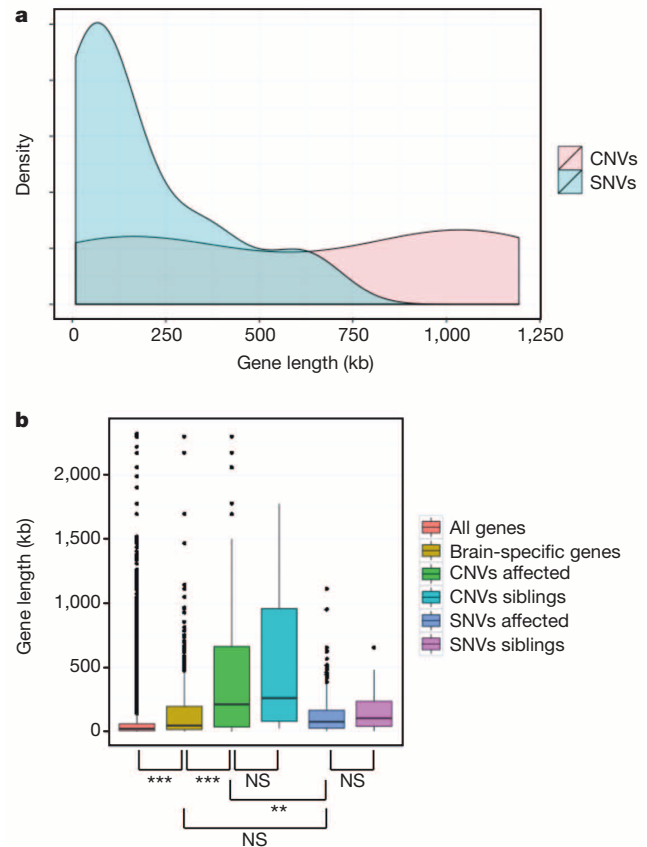


Figure 1 | Association between gene size and mutation types. **a**, Density plots of gene length for ASD genes with evidence for association according to the SFARI Gene database. Genes were divided to two groups on the basis of the type of mutations that implicated the gene in ASD, CNVs or SNVs. **b**, The distribution of gene length is presented by box plots for all genes in the genome, brain-specific genes and genes with *de novo* SNVs or CNVs identified by recent genome-wide studies. NS, non-significant; ** $P < 0.01$; *** $P < 0.001$.

King *et al.*¹, and as was previously reported⁹, we found that brain-specific genes are significantly larger than the average gene in the genome ($P = 6 \times 10^{-22}$). Whereas genes identified to be disrupted by *de novo* SNVs in ASD had a similar distribution of sizes as brain-specific genes ($P = 1$), the size of genes with *de novo* CNVs were significantly larger than either group ($P < 3 \times 10^{-3}$, $P < 5 \times 10^{-5}$, respectively). Furthermore, there was no difference in gene size between affected versus unaffected children for both *de novo* CNVs and SNVs ($P = 1$).

In summary, our analysis suggests that the association between large genes and ASD that was observed by King *et al.*¹ stems mainly from the method of implicating genes based on CNVs, and is not an inherent property of ASD risk genes.

Methods

The list of ASD genes was constructed based on the SFARI Gene database (accessed 11 December 2013). We discarded genes with no or minimal support for association (score > 3). Because of the difficulties to replicate genetic association and linkage results of ASD, candidate genes were considered only if the evidence was based on rare variants. The length of each gene was determined based on the largest transcript in the refSeq table of the UCSC Genome Browser (hg19 assembly). Genes were divided into CNVs or SNVs groups on the basis of the majority of studies that associated

the gene with ASD. In addition, we studied the length of genes found to be disrupted by *de novo* SNVs (nonsense, frameshift or splice site mutations)^{5–8}, and single genes affected by *de novo* CNVs^{3,4}. To test for differences in the distribution of gene sizes in different groups we performed a Kruskal–Wallis rank sum test on all groups (using the `kruskal.test` function in The R Project for Statistical Computing), followed by a pair-wise Mann–Whitney–Wilcoxon Test (using the `wilcox.test` function in R). *P* values were adjusted for multiple tests using the Bonferroni correction.

Shahar Shohat¹ & Sagiv Shifman¹

¹Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
email: sagiv@vms.huji.ac.il

Received 17 January; accepted 23 May 2014.

1. King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58–62 (2013).
2. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).

Zylka *et al.* reply

REPLYING TO S. Shohat & S. Shifman *Nature* **512**, <http://dx.doi.org/10.1038/nature13583> (2014)

Shohat and Shifman's analysis¹ indicates that long autism spectrum disorder (ASD) genes are overrepresented in the SFARI Gene/AutDB database (as of 11 December 2013) owing to the discovery method. We agree with their analysis and with the need to consider the strength of evidence behind each candidate gene. When our study was underway², SFARI Gene provided the only comprehensive list of autism candidate genes with confidence values. Subsequent to our publication, more genes have been added to this database and scored, highlighting the rapid pace of advances in the ASD field and the changing confidence behind each ASD gene.

We agree with Shohat and Shifman that the proportion of ASD genes that are long may drop as more ASD genes are identified. We did not account for how the discovery method used to identify a given ASD candidate, be it based on copy number variant (CNV) or single nucleotide variant (SNV), might affect average gene length in our study. However, it is also undeniable that many long genes are considered candidates in ASD pathology, such as *NRXN1* and *CNTNAP2*. Moreover, our mechanistic findings are not in dispute. Indeed, three other groups came to the same conclusion as us—that topoisomerases preferentially facilitate expression of long genes^{3–5}. Our study demonstrates an essential role for topoisomerases in transcriptional elongation of long neuronal genes and suggests a critical role for these enzymes in neurodevelopmental disorders like autism.

Shohat and Shifman¹ also suggest that the SFARI Gene database contains many genes with weak links to ASD pathology. In our study, we did not rank genes as stronger or weaker ASD candidates, and treated all equally. However, when the degree of evidence behind each candidate is taken into account, using the gene scoring module in SFARI Gene 2.0⁶ (as of 1 April 2014), it remains clear that numerous strong ASD candidate genes are very long (>200 kilobases). Thus, we feel our conclusion linking topoisomerases and gene length with autism is still warranted, but this remains to be tested more rigorously pending *in vivo* studies with animal models and additional human genetic studies.

Future studies are likely to validate additional long genes as strong ASD candidates. For example, *NRXN3* and *CNTN5* (1.5 and 1.3 megabase,

3. Levy, D. *et al.* Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886–897 (2011).
4. Sanders, S. J. *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
5. Iossifov, I. *et al.* *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
6. Neale, B. M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
7. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
8. Sanders, S. J. *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
9. Ben-David, E. *et al.* Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. *Hum. Mol. Genet.* **20**, 3632–3641 (2011).

Author Contributions Both authors conceived the project, analysed the data and wrote the paper.

Competing Financial Interests Declared none.

doi:10.1038/nature13583

respectively) are not yet scored in SFARI Gene, yet these genes are deleted in patients with ASD^{7,8} and both are significantly reduced in topotecan-treated neurons².

Ultimately, we agree that making conclusions about the nature of ASD genes is complicated by factors like the ones Shohat and Shifman describe¹ as well as by the evolving knowledge of autism genetics. Regardless, we identified a transcriptional mechanism that affects the expression of long genes, a number of which are currently classified as strong ASD candidates.

Mark J. Zylka^{1,2,3}, Ben D. Philpot^{1,2,3} & Ian F. King¹

¹Department of Cell Biology and Physiology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

²Carolina Institute for Developmental Disabilities, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

³UNC Neuroscience Center, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

email: zylka@med.unc.edu

1. Shohat, S. & Shifman, S. Bias towards large genes in autism. *Nature* **512**, <http://dx.doi.org/10.1038/nature13583> (2014).
2. King, I. F. *et al.* Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501**, 58–62 (2013).
3. Solier, S. *et al.* Transcription poisoning by topoisomerase I is controlled by gene length, splice sites, and miR-142–3p. *Cancer Res.* **73**, 4830–4839 (2013).
4. Teves, S. S. & Henikoff, S. Transcription-generated torsional stress destabilizes nucleosomes. *Nature Struct. Mol. Biol.* **21**, 88–94 (2014).
5. Veloso, A. *et al.* Genome-wide transcriptional effects of the anti-cancer agent camptothecin. *PLoS ONE* **8**, e78190 (2013).
6. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol. Autism* **4**, 36 (2013).
7. Vaags, A. K. *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am. J. Hum. Genet.* **90**, 133–141 (2012).
8. van Daalen, E. *et al.* Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. *Neurogenetics* **12**, 315–323 (2011).

doi:10.1038/nature13584

Seeds of selective nanotube growth

'Seed' molecules have been made that enable synthesis of just one kind of single-walled carbon nanotube, rather than a mixture of species. This paves the way for the preparation of pure samples of any nanotube species. [SEE LETTER P.61](#)

JAMES M. TOUR

The wonderful thing about single-walled carbon nanotubes (SWCNTs) is that more than 100 species can be produced by various growth methods. But this is also the most frustrating thing about them. It is expected that different nanotube species will have different applications, but the nanotubes generally form as mixtures of around 5–50 species from any preparation method^{1,2}. Moreover, separation methods are cumbersome because of the many species that can form. After many attempts over the past two decades to grow single species of SWCNTs, Sanchez-Valencia *et al.*³ present a route to success on page 61 of this issue.

Each SWCNT species can be defined by a pair of integers (n, m) called chirality indices, which describe how a graphene sheet (a single layer of carbon atoms in graphite) would hypothetically be rolled up to generate a tubular structure⁴. Chirality indices can be used to determine the two unique, fundamental parameters of each rolled-up graphene structure — the tube diameter and the angle with respect to a plane perpendicular to the tube's long axis at which the graphene would be rolled into a tube. The term 'chiral' is sometimes a misnomer, because chirality is a property associated with asymmetry, but some SWCNTs are not asymmetric.

Although there are many species of SWCNT (Fig. 1), there are only two main types: metallic nanotubes, which conduct electricity in the same way as gold or aluminium; and semiconducting nanotubes, whose electrical conductivities are tunable, as in the semiconductors silicon and gallium arsenide. Conductivities are determined by a property called the bandgap — the smaller the bandgap, the larger the room-temperature conductivity. Metallic SWCNTs have a bandgap of 0 electronvolts (eV), whereas semiconducting nanotubes have a bandgap that can vary from approximately 1 meV to 1.5 eV (ref. 5). Specific bandgaps are required for particular applications. For example, a bandgap of 0 eV is desirable for electrical wire and cable applications, whereas a larger bandgap is preferred for transistors. For photonic applications, different bandgaps are required to generate or detect different colours⁶.

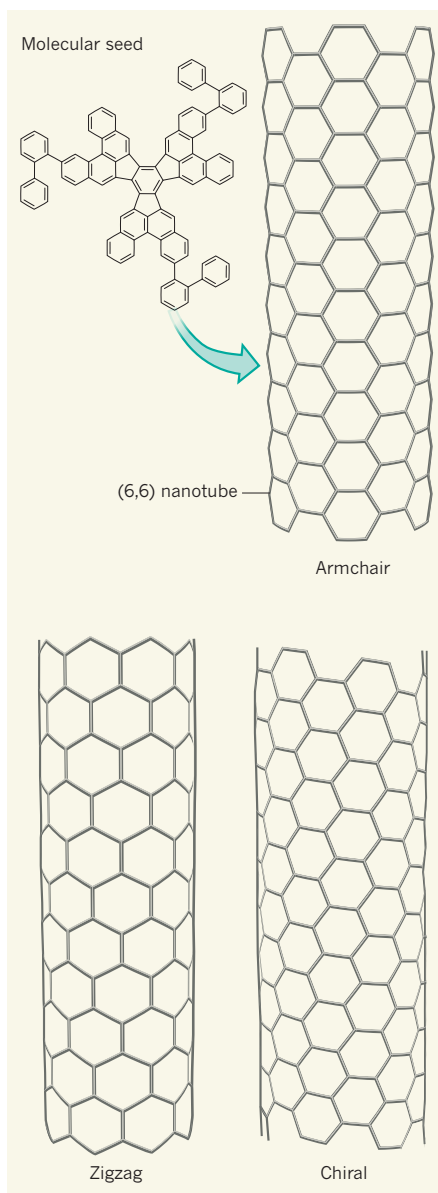


Figure 1 | Structural diversity of single-walled carbon nanotubes (SWCNTs). The orientations of the hexagonal rings of atoms in SWCNTs form the basis of three classes of nanotube, examples of which are shown here. Sanchez-Valencia *et al.*³ report a method for making an 'armchair' variety of SWCNT known as a (6,6) tube, starting from molecular 'seeds' on a catalytic platinum surface. The tubes form as a single product without contamination from zigzag or chiral nanotubes.

Sanchez-Valencia and co-workers prepared exclusively (6,6) species of tubes starting from predefined 'seeds' — organic molecules prepared by multi-step synthesis. They grew SWCNTs from each of these seeds on a platinum surface at 500 °C, using ethanol as a source of carbon atoms. The idea of using molecules to control the chirality of nanotubes is not new⁷, but the authors have taken the concept of seed design for specific nanotube growth to an extraordinary level: the precise arrangement of the atoms in the seed predefines the species of tube grown. Their work suggests that it should be possible to design and synthesize seeds for any desired SWCNT species.

Impressively, the researchers used scanning tunnelling microscopy to image the orientation of the seeds on the platinum surface and to take snapshots during the main phases of the nanotube-forming process — that is, the formation of a bowl-shaped cap from the seed and the subsequent 'base-growth' stages (in which the catalytic platinum atoms stay at the substrate surface and the top of the nanotube is catalyst-free). They also studied their nanotubes using Raman spectroscopy, observing a single peak in a band of the Raman spectrum that is diagnostic of the species of nanotube being analysed. This provided beautifully simple confirmation that only one species of nanotube forms from the seeds, and unambiguously pinpointed the nanotube structure. Furthermore, the authors performed extensive computational modelling to understand the different phases of the nanotube-forming process.

Sanchez-Valencia and co-workers' method is currently the only one that allows predictable control of the chirality of SWCNTs. In another approach reported⁸ this year, (12,6) SWCNTs were prepared as 92% of the nanotube mixture using a solid alloy catalyst, but the tube species that grew could not be predetermined. The relatively low temperature (500 °C) used in Sanchez-Valencia and colleagues' procedure probably helps to maintain species specificity because, at higher temperatures, small variations in the growth temperature can cause changes in the chirality index along a tube⁹.

Some might view the need for a 10-step organic synthesis of the seeds as an

overburdening limitation of the new approach. It is not. Consider that 1 mole of seeds is 6×10^{23} molecules, equating to 1.2 kilograms of material — a quantity that could easily be prepared by a chemical company. If, as Sanchez-Valencia *et al.* show, 50% of those seeds adopt the necessary conformation for growth at the platinum surface, then more than 5 tonnes of 10-micrometre-long SWCNTs could be obtained from just 1 mole of seeds.

However, further challenges remain. The new method produces nanotubes that stand perpendicular to the growth surface, like the bristles of a carpet. This minimizes entanglement of the nanotubes, but they will still form bundles when they reach a certain length. Many applications need SWCNTs to be unbundled, and so the nanotubes will need to be subsequently treated with solvent or wrapped with polymers. Furthermore, the surface area covered by SWCNTs using typical growth

methods is of the order of 1% (ref. 10); using Sanchez-Valencia and colleagues' approach, about 30 km² of platinum would be needed to accommodate 1 kg of seeds at this surface density, assuming that half of them grow. Placing nanotubes in arrays precisely where they are needed has also been a persistent obstacle to the development of many devices. Lastly, it remains to be seen whether molecular seeds can be made that selectively control the formation of other nanotube chiralities.

Sanchez-Valencia and colleagues' work represents a stellar breakthrough in the synthesis of SWCNTs. To those who have worked in this field for the past two decades, it is humbling to think that the selective growth of these diminutive structures has taken so long. But it is comforting to see it done so definitively. ■

James M. Tour is in the Departments of Chemistry and of Materials Science and

NanoEngineering, and at the Smalley Institute for Nanoscale Science and Technology, Rice University, Houston, Texas 77005, USA.
e-mail: tour@rice.edu

1. Kitiyanan, B., Alvarez, W. E., Harwell, J. H. & Resasco, D. E. *Chem. Phys. Lett.* **317**, 497–503 (2000).
2. Bachillo, S. M. *et al. Science* **298**, 2361–2366 (2002).
3. Sanchez-Valencia, J. R. *et al. Nature* **512**, 61–64 (2014).
4. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. in *Science of Fullerenes and Carbon Nanotubes* (Academic, 1996).
5. Matsuda, Y., Tahir-Kheli, J. & Goddard, W. A. III *J. Phys. Chem. Lett.* **1**, 2946–2950 (2010).
6. Hároz, E. H. *et al. Nanoscale* **5**, 1411–1439 (2013).
7. Fort, E. H. & Scott, L. T. *J. Mater. Chem.* **21**, 1373–1381 (2011).
8. Yang, F. *et al. Nature* **510**, 522–524 (2014).
9. Doorn, S. K. *et al. Phys. Rev. Lett.* **94**, 016802 (2005).
10. Pint, C. L., Xu, Y.-Q., Pasquali, M. & Hauge, R. H. *ACS Nano* **2**, 1871–1878 (2008).

CANCER

Directions for the drivers

A comparison of colorectal cancer and normal cells from 103 patients identifies dozens of genes that are differently expressed in tumour cells as a result of altered regulation of transcription. SEE LETTER P.87

GREG GIBSON

How important for cancer incidence and progression is genetic variation that affects gene expression? This fundamental question has received remarkably little attention in recent studies of cancer genomes¹, perhaps because of a prevailing view that the cancer-causing mutations that can be targeted by drugs are those that disrupt protein structure². On page 87 of this issue, Ongen *et al.*³ demonstrate how simultaneous gene-expression profiling and whole-genome genotyping can be used to dissect the regulation of gene transcription in colorectal cancer³. The findings provide two thought-provoking insights: that cancer-driving changes may be identifiable among an excess of regulatory mutations, and that 'cryptic' regulatory genetic variation has the potential to modify cancer progression.

It is well established that gene expression is altered in cancer. Despite their independent derivation, tumours of the same type tend to converge on a common, new gene-expression profile. Various studies, primarily from The Cancer Genome Atlas project¹, have noted differential transcription of tumour-driving and tumour-suppressing genes in advanced

tumours, but so many gene transcripts are altered in these tumours that it is difficult to know which ones 'drive' the altered behaviour and which ones are 'passengers', just going along for the ride. Furthermore, epigenetic alterations — those that modify gene expression without involving sequence mutations — have been implicated in cancer, including colorectal cancer^{4,5}. Broad surveys of transcriptional and epigenetic changes in

tumours have been conducted^{6,7}, but not on the scale and resolution achieved by Ongen and colleagues. They used a method known as RNA-Seq, in which the transcriptome of a cell (its whole complement of RNA molecules) is sequenced.

Over the past couple of years, sequencing of the exome of cancer cells (in essence, just the protein-coding regions) has suggested that around 250 genes are mutated in cancer cells significantly more often than expected by chance^{8–10}. Many of these are pan-cancer genes, and some are tumour-type specific. It is less straightforward to perform similar analyses for regulatory DNA, for two reasons: we are just beginning to learn how to identify regulatory functions in the hundreds of kilobases that surround genes, and altered gene expression is often due to changes in genes located elsewhere in the genome. Ongen *et al.* overcome these limitations by focusing on changes in the ratio of expression of heterozygous alleles (sites at which the DNA sequence differs between the two copies of the sequence in

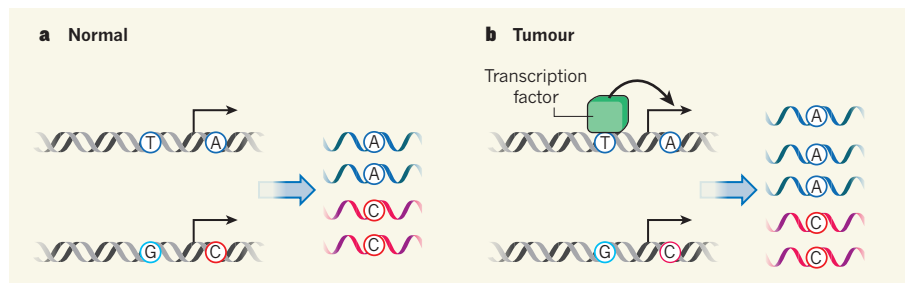


Figure 1 | Cryptic regulation in tumour cells. **a**, In normal cells, the two copies (alleles) of a gene in which the alleles are heterozygous, meaning that they contain a different nucleotide at a specific site (here, A and C), are both expressed at the same, low level. In this example, the alleles in the transcript are associated with a normally irrelevant, or 'cryptic', variant allele in a regulatory region of the gene (A with T and C with G). **b**, In a tumour cell that is characterized by enhanced expression of a transcription factor, this cryptic variant becomes relevant, because the transcription factor can bind to the T-containing allele but not to one containing G. This leads to enhanced transcription of the gene containing the A allele relative to that containing C, and thus an altered allelic-expression ratio.

a cell) between tumour and matched normal cells, as had also been done in another recent analysis of colorectal cancer¹¹. They call the hundreds of instances of this phenomenon that they find per sample 'genes with allelic dysregulation' (GADs).

Although allele-specific expression can also be attributed to changes at other genes, it is highly likely that in many cases it is due to a locally acting regulatory mutation. Ongen *et al.* observe a significant correlation between the somatic (non-germline) mutation rate and altered allele-specific expression and, in each of the 103 matched normal–tumour pairs they analysed, approximately 200 transcripts showed a cancer-specific deviation in the allelic expression ratio at heterozygous sites. Their interpretation is that one allele is transcribed more than the other owing to the action of regulatory-sequence variation.

The authors show that some of this deviation can be attributed to familiar cancer-associated mutation types, including loss of heterozygosity and copy-number alteration, and that some is due to inferred (yet to be defined) regulatory mutations. Tallying these instances over all of the samples, and taking two approaches to controlling for statistical biases, they arrive at a list of 71 GADs that occur more frequently in tumours than in normal cells, 9 of which are shared with an existing list of pan-cancer driver mutations⁸. These observations provide a smoking gun for the idea that regulatory mutations can drive cancer. Perhaps there is no need to distinguish them with their own name, but the term 'GPS mutations' comes to mind, because they are instructing driver mutations on what to do, but it is not altogether clear that the cancer cells would not still attain a tumorous state without their help — much like a satellite-based navigation system instructing a driver on how to get somewhere.

A related term, 'back-seat driver', has been invoked to describe another class of mutation that probably has a role in mediating cancer progression or metastatic spread, and that is conditional on the status of other driver mutations¹². Ongen and colleagues' second major contribution is to suggest that, in addition to GPS mutations in GADs, another important source of cancer regulation is cryptic genetic variation (Fig. 1). These are genetic variants that are not relevant under normal circumstances, but become so only in a perturbed state¹³. They may play a key part in modifying the expression of cancer-driving genes. Specifically, when the authors looked for common variants that associate with differences in gene expression among individuals, they found that at least one-third of the expression-regulating variants (eSNPs) are tumour specific.

Furthermore, these polymorphisms are enriched for binding sites for six known cancer-related transcription factors, all of which are upregulated in colorectal cancer. The idea is that when one of these factors (IRX3,

E2F4, NFIL3, TFAP2A, CUX1 or LEF1) is in excess, polymorphisms that in normal cells do not influence the expression of the adjacent gene become relevant. Whether or not these are key back-seat modifiers of cancer progression is unclear, because there is only a mild enrichment for these polymorphisms in a genome-wide association study (GWAS) of colorectal cancers¹⁴. Perhaps they would be more enriched in a GWAS that assessed tumour progression.

Two words of caution about this study are in order. The first is that there has been no attempt to demonstrate functionality of any of the candidate mutations — the findings are all based on statistical association. The second is that anyone using RNA-Seq quickly realizes that there are many points at which the analysis can provide divergent results. Given recent interest in the repeatability of findings, it could be argued that it would be a good idea for journals to require independent parallel analyses from a different group, conducted blind, to corroborate such results before publication. This suggestion is not made to denigrate the careful and insightful analyses reported by Ongen *et al.*, but is rather a generic comment on the inherent complexity of RNA-Seq, GWASs and enrichment analyses. Different analysts are likely to find quite different details. Yet the prospect that acquired variants drive cancer by controlling gene expression against a background of cryptic regulatory

modifiers opens up a new perspective on cancer research. Similar analyses can now be performed on other data sets, and also on diseases other than cancer in which regulation of gene expression is altered^{15,16}. The next challenge is to establish the clinical utility of the identified regulatory variation. ■

Greg Gibson is in the School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30309, USA.
e-mail: greg.gibson@biology.gatech.edu

1. The Cancer Genome Atlas Research Network. *Nature Genet.* **45**, 1113–1120 (2013).
2. Hopkins, A. L. & Groom, C. R. *Nature Rev. Drug Discov.* **1**, 727–730 (2002).
3. Ongen, H. *et al. Nature* **512**, 87–90 (2014).
4. Bogaert, J. & Prenen, H. *Ann. Gastroenterol.* **27**, 9–14 (2014).
5. The Cancer Genome Atlas Research Network. *Nature* **487**, 330–337 (2012).
6. Li, Q. *et al. Cell* **152**, 633–641 (2013).
7. Bell, J. T. *et al. Genome Biol.* **12**, R10 (2011).
8. Tamborero, D. *et al. Sci. Rep.* **3**, 2650 (2013).
9. Ciriello, G. *et al. Nature Genet.* **45**, 1127–1133 (2013).
10. Watson, I. R., Takahashi, K., Futreal, P. A. & Chin, L. *Nature Rev. Genet.* **14**, 703–718 (2013).
11. Lee, R. D.-W., Song, M.-Y. & Lee, J.-K. *Gene* **512**, 16–22 (2013).
12. Futreal, A. P. *Cancer Cell* **12**, 493–494 (2007).
13. Paaby, A. B. & Rockman, M. V. *Nature Rev. Genet.* **15**, 247–258 (2014).
14. Peters, U. *et al. Hum. Genet.* **131**, 217–234 (2012).
15. Barreiro, L. B. *et al. Proc. Natl Acad. Sci. USA* **109**, 1204–1209 (2012).
16. Kim, J. *et al. Genome Med.* **6**, 40 (2014).

This article was published online on 23 July 2014.

NEURODEGENERATION

Alzheimer's disease under strain

Two studies of amyloid- β protein aggregates, which cause Alzheimer's disease, find that different conformations of the aggregates can define different strains of the disorder, drawing parallels with prion diseases.

ADRIANO AGUZZI

A crucial event in the initiation of Alzheimer's disease is the formation of amyloid- β protein aggregates in the brain. Several studies indicate that these aggregates can acquire distinct shapes¹. Now, two papers^{2,3} published in *Proceedings of the National Academy of Sciences* demonstrate that each amyloid- β protein of a given shape can seed more aggregates of that same shape. This suggests that there are several variants of Alzheimer's disease, similar to influenza strains causing clinically distinct epidemics.

The prion, which causes spongiform encephalopathies (neurodegenerative diseases such as Creutzfeldt–Jakob disease in humans

and scrapie in sheep), remained shrouded in mystery for decades. But over the past 20 years, the 'protein-only' prion hypothesis⁴ has become widely accepted. This states that prions arise from an abnormal version of the cell-membrane protein PrP, called PrP^{Sc}, which forms misfolded protein aggregates, or plaques.

Surprisingly, given the protein-only hypothesis, prions isolated from different sources (such as sheep, deer and cattle) target different brain areas, induce unique symptoms and have varied incubation times, indicating that there are different strains⁵. More disconcertingly, strain-specific properties are maintained even when prions from different sources are transmitted through several generations of inbred mice with identical genes encoding

PrP (ref. 6; Fig. 1a). In typical viral and bacterial infections, genetic variations control strain-specific properties, but this mouse experiment suggests that the heritable traits of prions must somehow be encoded within the protein, most probably within its shape. Much evidence has accrued in favour of this theory, with several papers^{7–9} indicating that PrP^{Sc} can adopt different conformations.

Although the detailed structure of PrP^{Sc} remains unresolved, the protein seems to be very similar to amyloids, a family of filamentous structures that form protein aggregates, as PrP^{Sc} does. Amyloid aggregates are implicated in many neurodegenerative diseases — including Alzheimer's disease and Parkinson's disease — and are found in a growing list of body sites in addition to the brain. Therefore, understanding the details of prion and amyloid assembly and replication is of broad relevance.

If prions can self-assemble in different conformations, and amyloid- β shares properties with PrP^{Sc}, might amyloid- β also arise in different strains? To investigate this, Watts *et al.*² isolated amyloid- β from two patients who had died from the sporadic form of Alzheimer's disease, and from two patients with hereditary Arctic or Swedish forms of Alzheimer's disease. The mutation that causes Arctic Alzheimer's disease changes the DNA sequence of the gene encoding amyloid- β , whereas the Swedish mutation occurs outside this genetic region and augments the cleavage of amyloid- β from a larger precursor protein.

Watts and colleagues inoculated mice engineered to overexpress amyloid- β with brain samples from the four patients and from one healthy human control. Confirming previous studies¹⁰, they found that inoculation with amyloid- β -containing samples accelerated the formation of amyloid- β plaques in all mice. However, Swedish and sporadic amyloid- β differed from Arctic amyloid- β in many ways. In particular, the amyloid- β aggregates that formed in mice exposed to the Arctic inoculum were much less resistant to physical disassembly than were aggregates derived from any other inoculum. This may seem counterintuitive, because the Arctic mutation is highly aggressive, but it conforms with the finding¹¹ that more-fragile fibrils have greater seeding potential and, by inference, are more aggressive.

Crucially, the authors found that the Arctic inoculum maintained its strain-specific properties ('strainness') even when they recovered samples from inoculated mice and transmitted them to a second generation of genetically identical mice (Fig. 1b). The significance of this result may be clear to virologists, but such experiments have rarely, if ever, been performed in the context of non-PrP^{Sc} amyloids. Because the amyloid- β aggregates that are passed

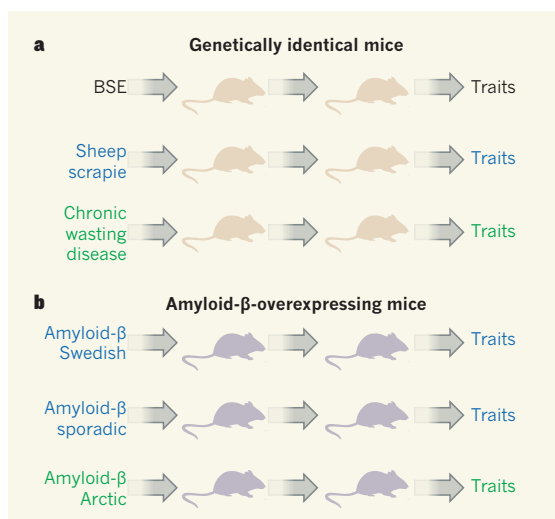


Figure 1 | Information encoded in protein aggregates. **a**, Prion diseases with differing characteristics, including bovine spongiform encephalopathy (BSE), sheep scrapie and chronic wasting disease, are caused by aggregation of the same protein, PrP^{Sc}. These distinct traits are maintained when disease samples are inoculated into the brains of mice and the accumulating aggregates are subsequently transferred through genetically identical mice. **b**, Similarly, Watts *et al.*² report that the strain-specific traits of three forms of Alzheimer's disease — Swedish and sporadic, which have similar traits, and Arctic, which is more aggressive — are maintained during inoculation and serial transmission of each strain through two generations of genetically identical mice that overexpress amyloid- β . Stöhr *et al.* produce similar results with two synthetic forms of amyloid- β (not shown). This suggests that strain-specific traits of prion diseases and Alzheimer's disease are encoded by the conformations in which protein aggregates form.

onto the second mouse are almost entirely of mouse origin, any retained strainness must be enciphered in the conformation of amyloid- β protein, rather than in its gene sequence.

In an accompanying study, Stöhr *et al.*³ provide evidence that two synthetic variants of amyloid- β that are 40 or 42 amino-acid residues in length ($A\beta_{40}$ and $A\beta_{42}$) cause distinct disease characteristics when injected into mice overexpressing amyloid- β . This finding suggests that distinct amyloid- β strains can be generated from chemically defined, homogeneous proteins — or that $A\beta_{40}$ and $A\beta_{42}$ cause different disease characteristics because they have different gene sequences that dictate different kinetics of aggregate assembly¹².

The current studies present convincing evidence for the existence of conformation-enciphered amyloid- β strains in patients with Alzheimer's disease, but do not provide structural information about the protein conformations involved. Last year, conformations of amyloid- β in two patients with sporadic Alzheimer's disease were studied¹ at extremely high resolution. This experiment showed that the two patients harboured distinct conformations of amyloid- β aggregates, potentially representing distinct strains.

Excitingly, the existence of amyloid- β strains might explain some of the current limitations

of therapies for Alzheimer's disease. Immunization with amyloid- β antibodies is extremely effective in mouse models of Alzheimer's disease¹³, but translation to humans has proved disappointing. Might distinct amyloid- β strains be better targeted with strain-specific antibodies? If so, precise strain testing in Alzheimer's disease may lead to better patient stratification and higher response rates¹⁴.

These papers assert that amyloid- β aggregates are prions. However, key differences set PrP^{Sc} apart from amyloids. Prions are by definition transmissible, and have caused epidemics — bovine spongiform encephalopathy, for example, has killed almost 200,000 cattle, and its transmission to humans in the form of variant Creutzfeldt–Jakob disease has caused more than 200 deaths. By contrast, there is no evidence that Alzheimer's disease is infectious among humans, and the transmissibility of amyloid- β is limited to mice that are genetically programmed to be aggregation-prone. Therefore, amyloid- β aggregates are distinct from prions — even if they exhibit strainness. These clinical and epidemiological differences are important for human medicine, and are evidence against lumping amyloidoses and prion diseases together. In fact, it might be more appropriate to classify amyloid- β aggregates as prionoids¹⁵.

Moreover, if we classify amyloid- β aggregates as prions, we need to treat them accordingly. PrP^{Sc} is subjected to mandatory confinement (biosafety level 2 or 3, depending on the prion strain), and in the United States, prions are considered to be bioterrorism agents. Expensive medical devices that have been used on people who are infected with prions must typically be destroyed to prevent accidental disease transmission. If amyloid- β aggregates are truly equivalent to prions, such draconian measures may be necessary — and inevitable. But it would seem prudent to accrue more evidence that amyloid- β is genuinely infective before implementing such measures. ■

Adriano Aguzzi is at the Institute of Neuropathology, University Hospital Zurich, Zurich CH-8091, Switzerland.
e-mail: adriano.aguzzi@usz.ch

1. Lu, J. X. *et al.* *Cell* **154**, 1257–1268 (2013).
2. Watts, J. C. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 10323–10328 (2014).
3. Stöhr, J. *et al.* *Proc. Natl Acad. Sci. USA* **111**, 10329–10334 (2014).
4. Prusiner, S. B. *Science* **216**, 136–144 (1982).
5. Fraser, H. & Dickinson, A. G. *J. Comp. Pathol.* **83**, 29–40 (1973).
6. Sigurdson, C. J. *et al.* *J. Virol.* **80**, 12303–12311 (2006).
7. Peretz, D. *et al.* *Neuron* **34**, 921–932 (2002).
8. Safar, J. *et al.* *Nature Med.* **4**, 1157–1165 (1998).
9. Sigurdson, C. J. *et al.* *Nature Meth.* **4**, 1023–1030 (2007).

10. Meyer-Luehmann, M. *et al. Science* **313**, 1781–1784 (2006).
 11. Knowles, T. P. *et al. Science* **326**, 1533–1537 (2009).
 12. Meisl, G. *et al. Proc. Natl Acad. Sci. USA* **111**, 9384–9389 (2014).
 13. Schenk, D. *et al. Nature* **400**, 173–177 (1999).
 14. Aguzzi, A. & Gitler, A. D. *Cell* **154**, 1182–1184 (2013).
 15. Aguzzi, A. *Nature* **459**, 924–925 (2009).

ASTROPHYSICS

Portrait of a doomed star

Some stars explode in thermonuclear supernovae, but understanding of why this occurs comes mainly from indirect clues. Now, the progenitor of a member of a strange class of such explosions may have been detected directly. [SEE LETTER P.54](#)

STEPHEN JUSTHAM

Explaining the nature of supernovae is one of the classical problems in astronomy. Supernovae are not only enticing mysteries in which explosions of awesome power and brilliance are perpetrated by well-hidden culprits, they are also of broad importance to astrophysics — so our lack of certainty about the progenitors of some supernovae is embarrassing. On page 54 of this issue, McCully *et al.*¹ report that a combination of good fortune and careful analysis has pointed them to the probable system behind a particularly puzzling type of supernova.

In ancient times, people interpreted new lights in the sky as heavenly signs of earthly fates. Now, astronomers use a form of supernova known as type Ia to infer the history and future of the Universe. Those ancient new lights appeared to be new stars only because the stars that produced the celestial displays had previously been too faint to see. We now have a similar problem with supernovae. Although supernova explosions can be seen across more than half the age of the Universe,

detecting pre-explosion stars is difficult even when they are massive and luminous. For such supernovae we still only have a small number of definitive pre-explosion detections. The progenitors of type Ia supernovae have proved even more elusive. This type of supernova is thought to occur when a star called a white dwarf undergoes runaway nuclear fusion, but we have yet to directly see what circumstances cause this to happen. One such system may have been detected in X-rays², but perhaps the clearest trace of these progenitors has been in material lingering close to the explosion^{3–5}. However, theorists have been more successful in making multiple models consistent with the presence of that material than in convincing everyone to agree on one interpretation.

We are therefore lucky that two supernovae separated by only a decade have been detected in the spiral galaxy NGC 1309 (Fig. 1). After the first explosion, detected in 2002, a team led by one of McCully's co-authors used the Hubble Space Telescope to stare at the host galaxy⁶. The astronomers did this so that they could compare the distance to NGC 1309

inferred from two methods — one using the 2002 supernova and the other involving a type of star for which the method has been well-calibrated with nearby examples — and improve the precision with which we measure even larger cosmological distances. So when the second explosion occurred, in 2012, there was a wealth of data to examine for signs of the supernova progenitor.

At first, the 2012 supernova was classified as a type Ia (ref. 7), although one with very unusual features. However, it and similar supernovae have become known as type Iax, following the name of a peculiar prototype event called SN 2002cx (refs 8,9). Astronomers sometimes worry too much about classification: dividing objects that show continuous variation in physical properties such as luminosity and velocity into discrete categories can become unhelpful. However, this new name reflects the realization that the 2002cx-like explosions may be even less like normal type Ia supernovae than first thought. Superficially similar qualities might have previously misled us into doing something akin to classifying a platypus as a type of duck.

McCully and colleagues have found a source in the pre-explosion images from the Hubble Space Telescope whose location closely matches that of the 2012 supernova. They also show that this alignment has only a roughly 1% probability of being a coincidence, which leads them to conclude that this source is very likely to have produced the supernova. Although apparently exciting coincidences can happen by chance, there have been very few pre-supernova observations that could have identified a progenitor that resembles this one. So this result is not due to astronomers taking snapshot after snapshot and waiting for an interesting-looking but meaningless alignment to randomly occur.

Assuming that the observed source is the progenitor, the first question in interpreting the results is whether it was emitting mainly simple starlight. Because we might also expect light to be generated by matter falling onto the white dwarf before it explodes, it is possible that McCully *et al.* have detected a signal from this process. However, the authors argue that it is more likely that the source they have identified is a star made of helium, which has been transferring matter onto an unseen white dwarf. This option is tempting because it matches some models for type Iax supernovae in which the white dwarf accretes a surface helium layer that eventually ignites in a thermonuclear runaway, in turn triggering fusion in the underlying star.

If this explanation is true, then the source should still be present when the light from the supernova has faded sufficiently to allow the researchers to look for it again. However, the supernova might well have heated the helium star, stripped it of some surface layers or polluted it. It would be frustrating if the star

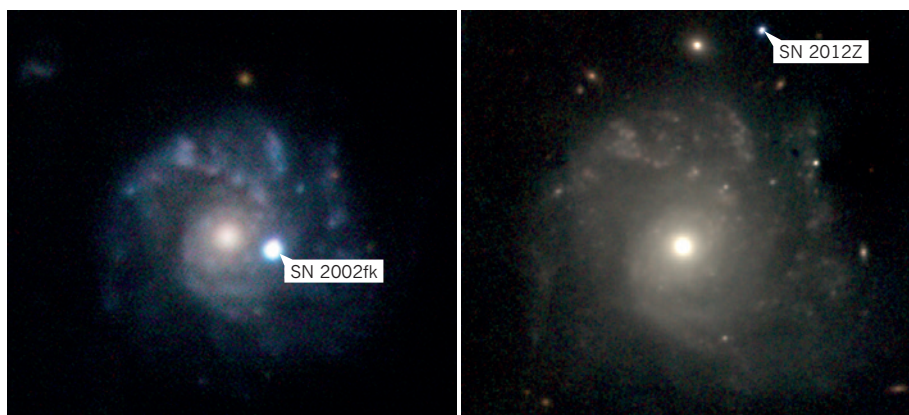


Figure 1 | A double bill. The spiral galaxy NGC 1309, which is about 30 million parsecs from Earth, hosted a type Ia supernova in 2002, SN 2002fk, and then a type Iax explosion in 2012, SN 2012Z. McCully *et al.*¹ have detected the probable progenitor of SN 2012Z in pre-explosion images obtained by the Hubble Space Telescope (see Fig. 1 of the paper¹).

W. LI & A. V. FILIPPENKO/UC, BERKELEY; G. H. MARION, S. W. JHA & R. P. KIRSHNER

appears exactly as it did before the explosion, and in that case we would have to worry more about whether the true progenitor was too faint to have been detected. It is to be hoped that future observations of this source will help us to understand not just type Ia events, but also the impact of supernovae on nearby stars. Another possibility noted by McCully and colleagues is that this supernova was the death of a massive star, in which case the source will have disappeared. This conclusion would show that Ia events truly are produced by qualitatively different systems from type Ia supernovae.

If McCully *et al.* have identified the progenitor, then the observation will be one of the most memorable signposts on the road

to understanding supernovae. And because it may improve our knowledge of what can happen after a layer of helium ignites on the surface of another potentially explosive collection of fusion fuel, it could also help to explain the progenitors of other recently discovered events for which variations of this model have been proposed^{10–13}. The 2012 supernova in NGC 1309 has not yet yielded all its secrets, but this discovery might help to solve mysteries both old and new. ■

Stephen Justham is at the National Astronomical Observatories, Chinese Academy of Sciences, 100012 Beijing, China. e-mail: sjustham@nao.cas.cn

1. McCully, C. *et al.* *Nature* **512**, 54–56 (2014).
2. Voss, R. & Nelemans, G. *Nature* **451**, 802–804 (2008).
3. Patat, F. *et al.* *Science* **317**, 924–926 (2007).
4. Sternberg, A. *et al.* *Science* **333**, 856–859 (2011).
5. Dilday, B. *et al.* *Science* **337**, 942–945 (2012).
6. Riess, A. G. *et al.* *Astrophys. J. Suppl. Ser.* **183**, 109–141 (2009).
7. Cenko, S. B. *et al.* *Cent. Bur. Electron. Telegrams* **3014**, 1 (2012).
8. Li, W. *et al.* *Publ. Astron. Soc. Pac.* **115**, 453–473 (2003).
9. Foley, R. J. *et al.* *Astrophys. J.* **767**, 57 (2013).
10. Perets, H. B. *et al.* *Nature* **465**, 322–325 (2010).
11. Kasliwal, M. M. *et al.* *Astrophys. J.* **755**, 161 (2012).
12. Poznanski, D. *et al.* *Science* **327**, 58–60 (2010).
13. Kasliwal, M. M. *et al.* *Astrophys. J. Lett.* **723**, L98–L102 (2010).

HIV

Early treatment may not be early enough

Giving monkeys antiretroviral therapy from just three days after exposure to simian immunodeficiency virus does not prevent a subsequent rebound of viral replication, suggesting that viral reservoirs are established early. [SEE LETTER P.74](#)

KAI DENG & ROBERT F. SILICIANO

Although antiretroviral therapy (ART) is successful in controlling HIV-1 replication, the virus persists in a stable latent reservoir in infected cells that have entered a resting state^{1–4}. In these immune cells, called resting memory CD4⁺ T cells, the viral genome hides as pure genetic information integrated into the cells' DNA (as proviral DNA), unaffected by ART or immune responses. But when the cells are subsequently activated, this viral reservoir again releases replication-competent HIV-1, and it is therefore considered the main barrier to curing HIV-1 infection⁵. Despite vigorous efforts to understand the latent reservoir in the hope of finding ways to purge it, it has been unclear when it is seeded and whether early treatment can prevent this. On page 74 of this issue, Whitney *et al.*⁶ provide evidence in the simian immunodeficiency virus (SIV) model of HIV-1 infection that the reservoir is seeded very early — during the first few days of infection.

It had been assumed that the initial seeding of the latent reservoir occurs during acute HIV-1 infection, at a point when viraemia — the presence of viruses in the blood — has risen to a high level⁷. It was proposed that if ART is begun before peak viraemia occurs, this might prevent the reservoir from becoming established, or at least significantly reduce its size. Recent clinical studies confirmed that early ART can indeed

reduce the size and dissemination of the viral reservoir^{8–10}, and even promote long-term control of the virus in some infected individuals after treatment has ended^{11,12}.

Interest in the possibility of viral-eradication strategies based on early ART initiation was further heightened by the case of the 'Mississippi baby', a child who was born to an infected mother and who had around 20,000 viral copies per millilitre of blood plasma at birth¹³. The child was started on ART 30 hours after delivery and was treated for 18 months. The virus was undetectable by 29 days, and remained so for 27 months after treatment was stopped, when rebound viraemia was detected¹⁴. Delayed and highly variable time to rebound is the predicted outcome of interventions that reduce the reservoir to a very low level¹⁵.

To evaluate the temporal dynamics of initial viral-reservoir seeding, Whitney and colleagues treated rhesus macaques starting

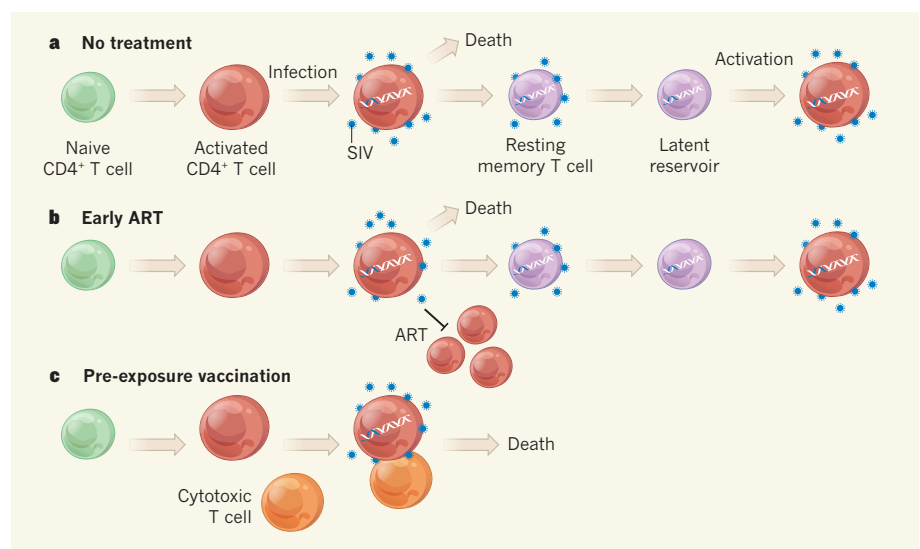


Figure 1 | SIV eradication strategies. **a**, Activation of naive CD4⁺ T cells renders the cells highly susceptible to infection with simian immunodeficiency virus (SIV), which becomes integrated into the host-cell genome to allow viral replication. Most CD4⁺ T cells die rapidly after infection, but a small fraction survives and reverts back to a resting memory state, in which SIV gene expression is turned off, resulting in a latent reservoir of the virus. Subsequent activation of these cells can restart virus production. **b**, Antiretroviral therapy (ART) soon after infection can stop more cells from becoming infected, but does not affect the fate of already infected cells, and some survive to seed the latent reservoir. Whitney *et al.*⁶ show that a viral reservoir is established within days of SIV infection. **c**, In vaccinated animals, SIV-specific cytotoxic T cells that are generated in response to the vaccine can kill infected cells before they revert back to the resting state, thereby preventing the establishment of a latent reservoir.

on days 3, 7, 10 or 14 after SIV infection. Although initiating treatment on days 7, 10 and 14 significantly reduced peak plasma virus levels, treatment from day 3 completely blocked the emergence of detectable viraemia; this was also evidenced by the absence of SIV-specific antibody-based and cellular immune responses in these animals. The authors found no proviral DNA in the animals' peripheral-blood mononuclear cells (which include CD4⁺ T cells) before treatment initiation on day 3, but proviral DNA was already detectable in their lymph nodes and the mucosal linings of the gastrointestinal tract. This crucial finding suggests that the viral reservoir may be first seeded in the lymphoid and mucosal tissues, a result with important implications for HIV-1 eradication strategies.

Most significantly, the authors observed viral rebound in all animals after ART was stopped. This occurred even when ART that fully suppressed detectable viraemia was initiated at day 3 and continued for 6 months, a treatment period that allows elimination of labile infected cells and thus reveals stable reservoirs. The observed rebound suggests that the viral reservoir is seeded surprisingly early in SIV-infected animals. However, the animals treated from day 3 showed a slightly delayed viral rebound compared with those that started ART at later times. Using a sophisticated model of viral dynamics, the authors show that the time to viral rebound is correlated with total viraemia during the acute phase of infection.

These data indicate that the viral reservoir could be seeded substantially earlier than previously assumed — a sobering finding that poses additional hurdles to HIV-1 eradication efforts. If this evidence from SIV-infected animals reflects what happens early in HIV-1 infection in humans, it would mean that it is nearly impossible to initiate ART before viral reservoirs have seeded, because viraemia is not detectable at this point. In other words, reservoir seeding precedes any clinical evidence of infection. However, although early treatment may not prevent reservoir seeding, it has been consistently shown to reduce the size of the latent reservoir^{8–10}, and infected individuals who receive early treatment could have a lower barrier to cure in future eradication strategies.

Whitney and colleagues' findings are of particular interest in light of a study last year¹⁶ reporting that a disseminated SIV infection could be cleared by vaccine-induced T-cell-based immune responses. The different outcomes of these two studies may be partly due to the fate of infected cells during acute infection. Early initiation of ART immediately stops subsequent new infection of susceptible cells, but does not affect the fate of cells that are already infected. A small fraction of these infected cells survive and revert back to a resting state, thereby seeding the latent reservoir (Fig. 1). By contrast, vaccinated animals have pre-existing

SIV-specific cytotoxic T cells that can clear the infected cells before they go into latency, thus preventing the viral reservoir from becoming established.

It remains to be seen whether clinical studies will confirm Whitney and colleagues' observations, because substantial differences exist between SIV infection in rhesus macaques and HIV-1 infection in humans. As mentioned by the authors, the SIV dose used in their study may be much higher than the typical amount of HIV-1 involved in sexual transmission, perhaps resulting in a higher level of early viral replication. Nevertheless, the striking findings of the early seeding of the viral reservoir in mucosal and lymphoid tissues before viraemia is detected suggest that new approaches, in addition to early treatment, will be necessary to eradicate HIV-1 infection. ■

Kai Deng and Robert F. Siliciano are in the Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. **R.F.S.** is also at the Howard Hughes Medical Institute, Baltimore.

e-mail: rsiliciano@jhmi.edu

1. Finzi, D. *et al.* *Science* **278**, 1295–1300 (1997).
2. Chun, T.-W. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 13193–13197 (1997).
3. Finzi, D. *et al.* *Nature Med.* **5**, 512–517 (1999).
4. Siliciano, J. D. *et al.* *Nature Med.* **9**, 727–728 (2003).
5. Richman, D. D. *et al.* *Science* **323**, 1304–1307 (2009).
6. Whitney, J. B. *et al.* *Nature* **512**, 74–77 (2014).
7. Chun, T.-W. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 8869–8873 (1998).
8. Chun, T.-W. *et al.* *J. Infect. Dis.* **195**, 1762–1764 (2007).
9. Hocqueloux, L. *et al.* *J. Antimicrob. Chemother.* **68**, 1169–1178 (2013).
10. Ananworanich, J. *et al.* *PLoS ONE* **7**, e33948 (2012).
11. Salgado, M. *et al.* *Retrovirology* **8**, 97 (2011).
12. Sáez-Cirión, A. *et al.* *PLoS Pathogens* **9**, e1003211 (2013).
13. Persaud, D. *et al.* *N. Engl. J. Med.* **369**, 1828–1835 (2013).
14. Ledford, H. *Nature* <http://dx.doi.org/10.1038/nature.2014.15535> (2014).
15. Hill, A. L., Rosenbloom, D. I. S., Fu, F., Nowak, M. A. & Siliciano, R. F. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1406663111> (2014).
16. Hansen, S. G. *et al.* *Nature* **502**, 100–104 (2013).

This article was published online on 20 July 2014.

EVOLUTION

Tooth structure re-engineered

Mice deficient in the EDA protein lack normal tooth features. Restoring EDA in embryonic teeth at increasing doses has now been found to recover these dental features in a stepwise pattern that mimics evolution. SEE ARTICLE P.44

ZHE-XI LUO

A fundamental connection between developmental changes and evolution has long been established¹. This link is gaining renewed emphasis² as molecular studies shed new light on evolution by revealing many genetic modifications that alter developmental processes, in turn changing an organism's shape and structure. On page 44 of this issue, Harjunmaa *et al.*³ report that, by simply tinkering with the genes and signalling pathways that control the shape of developing teeth, they have remade several different tooth structures *in vitro*. These structures draw a striking parallel with how teeth evolved from those of distant mammalian ancestors to the teeth of modern-day rodents.

Some lineages of therians (marsupial and placental mammals and their kin) that lived in the Mesozoic era, 252 million to 66 million years ago, had 'tribosphenic' molars⁴. The taller front end of the tribosphenic lower molar, called the trigonid, had three cusps — the raised points on the crown

of the tooth — for shearing food. The lower back end, the talonid, had a basin-like surface for grinding food^{5,6}. The trigonid and talonid of Mesozoic mammals are still recognizable in modern-day rodents, albeit in a highly modified form. As rodents arose from ancestral mammals and diversified into many lineages, cusps that were separate in the Palaeocene epoch 66 million to 56 million years ago⁷ became progressively connected by crests — which are more effective for chewing plant food — in a 'cusp-to-crest' dental evolution that occurred in many rodent groups^{8,9}.

The *ectodysplasin A* (*Eda*) gene encodes a vertebrate signalling protein that is involved in the development of a wide array of structures, from hair to sweat glands¹⁰. In the embryonic tooth, the EDA protein is active in enamel knots, which are signalling centres and the precursors of adult tooth structures. EDA regulates the position and size of future tooth cusps and their connecting crests⁸. Mice that do not express *Eda* lack normal cusps and crests, and instead have only basic, generalized teeth¹⁰.

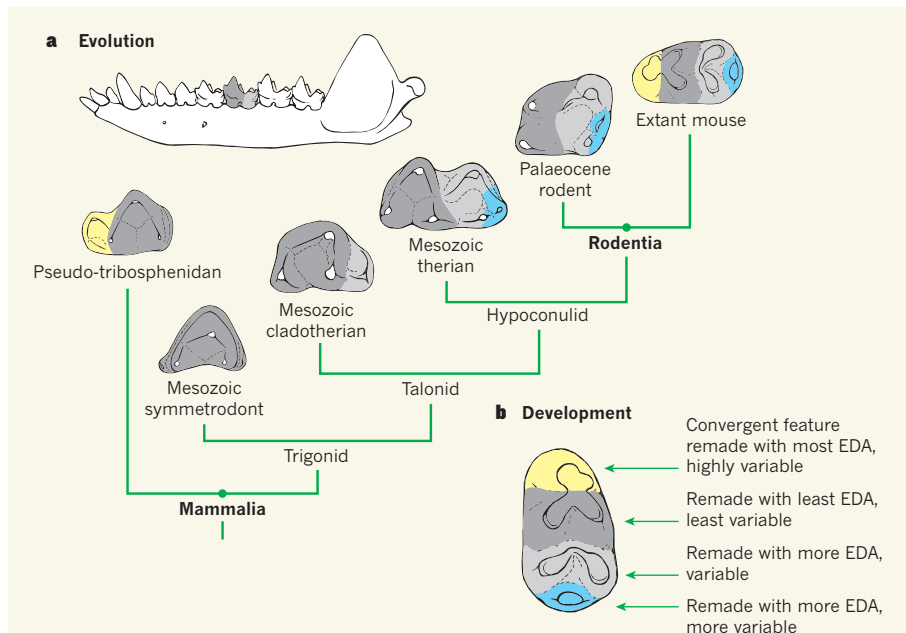


Figure 1 | Reconstructing tooth evolution *in vitro*. **a**, As mammals evolved, their molars (one indicated in the jaw) became ever more complex, because extra tooth features evolved over time. Structures called trigonids (dark grey) evolved first, in early mammals such as the Mesozoic symmetrodonts, followed by talonids (light grey) in a group of Mesozoic therians called cladotherians. Hypoconulids (blue) evolved in Mesozoic therians, and anteroconids (yellow) in advanced rodents. The anteroconids are similar in structure to pseudo-talonids, which evolved separately (convergently) in pseudo-tribosphenic teeth in an early-divergent clade of mammals (the pseudo-talonid is also indicated in yellow). **b**, Deletion of the *Eda* gene in mouse embryos results in the loss of all of these tooth features. Harjunmaa *et al.*³ show *in vitro* that addition of the EDA protein to embryonic teeth from *Eda*-deficient mice in increasing doses can replay the steps of evolution. Furthermore, features that evolved longer ago respond in a less variable manner than features that evolved more recently.

Harjunmaa and colleagues grew *Eda*-deficient teeth *in vitro*, and found that cusps and crests could be restored by adding EDA. They demonstrated, with the aid of computer models, that different doses of EDA alter tooth morphogenesis (the process by which structures are shaped as they develop), akin to the dental transformations that occurred as rodents evolved from their Mesozoic mammalian ancestors (Fig. 1). For example, the trigonid — the first part of the tribosphenic molar to have evolved — is regenerated with only a small dose of EDA. However, a higher dose of EDA is required to restore the talonid, which evolved more recently⁵.

The cusp-to-crest morphogenesis of mouse molars is controlled by a gene network that includes genes encoding the signalling proteins fibroblast growth factor 3 (*Fgf3*; ref. 9) and sonic hedgehog (*Shh*)¹¹. An increase or decrease in *Fgf3* causes over- or under-development of tooth features, respectively⁹. Harjunmaa and co-workers found that reducing the concentration of SHH in *Eda*-deficient teeth regenerated the ancestral features of Palaeocene rodents, reversing the cusp-to-crest transformation of modern-day rodents. Thus, molecular manipulations that alter tooth morphogenesis *in vitro* can replay evolution, either forward, to mimic the fossil record, or in reverse.

Perhaps the most exciting insight from

Harjunmaa and colleagues' work is that ancestral structures show a more uniform response to the addition or removal of EDA or SHH than those that evolved more recently or independently in different lineages (convergent evolution). For example, addition of a low dose of EDA reliably restored the trigonid, as expected of ancestral features, which are typically evolutionarily well conserved owing to their long history. By contrast, higher doses restored the talonid in many, but not all, teeth — development of this feature was more variable in response to EDA. This is consistent with the theory⁶ that the talonid basin evolved convergently in different mammal lineages, but has reduced in size in some carnivorous or insectivorous mammals.

The hypoconulid is a talonid cusp in some mammals, but is enlarged and forms a separate lobe in mice. The authors found that full development of this structure requires a higher dose of EDA than does the rest of the talonid, and shows even wider variation in its response to EDA. Finally, the anteroconid in mice — another feature that arose late in rodent evolution — requires the highest EDA dose to regenerate, and shows the broadest variation when regenerated. Its position on the tooth corresponds to the 'pseudo-talonid' that arose in some early-divergent mammals that died out before the end of the Mesozoic. Harjunmaa and co-workers' experiment

therefore demonstrates that modern-day mice still have the developmental potential to replicate evolutionary events that occurred long ago, in the now-extinct mammals of the Mesozoic¹².

The level of EDA required to give rise to individual molar characteristics therefore seems to provide information about how robust their development is. When studying how morphogenesis drives evolution¹³, it will be crucial to bear in mind that the sensitivity of a particular tooth feature to EDA activity may indicate the likelihood of an evolutionary transformation producing that feature. For example, as mentioned above, talonid-like features evolved twice — in the basal diversification of modern mammals and in early-divergent groups of the Mesozoic. Variable sensitivities to gene-expression dosage and signalling strength can serve as a measure of the evolutionary variability of each tooth feature, and may underpin the many convergences and reversals of tooth evolution observed in the mammalian fossil record.

Eda and *Shh* have varying effects on many vertebrate structures, so it can be hard to tease apart which evolutionary feature is controlled by which part of the gene network. Harjunmaa *et al.* have cleared this hurdle in a welcome development that brings us closer to being able to test how changes in morphogenesis affect the final shape of evolving teeth as seen in the fossil record. Genetic engineering of developmental processes *in vitro* is a fruitful way to decipher how the shape of organs or other biological structures is modified by evolution. ■

Zhe-Xi Luo is in the Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois 60637, USA.
e-mail: zxlue@uchicago.edu

- Gould, S. J. *Ontogeny and Phylogeny* (Belknap, 1985).
- Abzhanov, A. *Trends Genet.* **29**, 712–722 (2013).
- Harjunmaa, E. *et al.* *Nature* **512**, 44–48 (2014).
- Kielan-Jaworowska, Z., Cifelli, R. L. & Luo, Z.-X. *Mammals from the Age of Dinosaurs: Origins, Evolution, and Structure* (Columbia Univ. Press, 2004).
- Crompton, A. W. *Zool. J. Linn. Soc.* **50** (Suppl. 1), 65–87 (1971).
- Luo, Z.-X., Cifelli, R. L. & Kielan-Jaworowska, Z. *Nature* **409**, 53–57 (2001).
- Meng, J. & Wyss, A. R. *J. Mamm. Evol.* **8**, 1–71 (2007).
- Gomes-Rodrigues, H. *et al.* *Nature Commun.* **4**, 2504 (2013).
- Charles, C. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 22364–22368 (2009).
- Mikkola, M. J. & Thesleff, I. *Cytokine Growth Factor Rev.* **14**, 211–224 (2003).
- Cho, S.-W. *et al.* *Development* **138**, 1807–1816 (2011).
- Luo, Z.-X., Ji, Q. & Yuan, C.-X. *Nature* **450**, 93–97 (2007).
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* 2nd edn (Blackwell, 2005).

This article was published online on 30 July 2014.

Convergence of terrestrial plant production across global climate gradients

Sean T. Michalet¹, Dongliang Cheng², Andrew J. Kerkhoff³ & Brian J. Enquist^{1,4,5,6}

Variation in terrestrial net primary production (NPP) with climate is thought to originate from a direct influence of temperature and precipitation on plant metabolism. However, variation in NPP may also result from an indirect influence of climate by means of plant age, stand biomass, growing season length and local adaptation. To identify the relative importance of direct and indirect climate effects, we extend metabolic scaling theory to link hypothesized climate influences with NPP, and assess hypothesized relationships using a global compilation of ecosystem woody plant biomass and production data. Notably, age and biomass explained most of the variation in production whereas temperature and precipitation explained almost none, suggesting that climate indirectly (not directly) influences production. Furthermore, our theory shows that variation in NPP is characterized by a common scaling relationship, suggesting that global change models can incorporate the mechanisms governing this relationship to improve predictions of future ecosystem function.

Annual net primary production (NPP; $\text{g m}^{-2} \text{yr}^{-1}$) of woody plants is a major component of the terrestrial carbon cycle¹. Although it has long been known that NPP correlates with mean annual temperature^{2,3} and mean annual precipitation^{3–5}, efforts to understand the interactions between climate and metabolism have resulted in two conflicting generalizations. On the one hand, across a wide range of ecologically relevant temperatures, rates of photosynthesis and respiration increase approximately exponentially with temperature to a critical value beyond which rates decrease⁶. Furthermore, photosynthetic rate is limited by water availability, as reflected by a positive relationship between NPP and precipitation^{4,5}. These patterns suggest that climate influences NPP directly via metabolic kinetics. On the other hand, across broad environmental gradients, local adaptation of thermal and edaphic tolerances may dampen physiological responses^{7,8}. Under this scenario, correlations between NPP and annual climate variables result not directly from variation in metabolic rates, but rather indirectly via variation in plant size and stand biomass^{8–10}, stand age structure⁷ and growing season length¹⁰.

Metabolic scaling theory for NPP

To assess these differing viewpoints, we build upon metabolic scaling theory of forest structure and dynamics^{11,12} to test the relative importance of direct and indirect climate effects on NPP (Supplementary Information). First, according to metabolic scaling theory, variation in instantaneous rates of respiration, photosynthesis and growth scale predictably with plant size and temperature. Second, extensions of metabolic scaling theory to whole-ecosystem functioning predict that NPP will scale with stand biomass and size of the largest individual^{7,11–13}. Third, to incorporate the effects of temperature, precipitation, growing season length and plant age, we assume that their effects are inherently multiplicative¹⁴. As a result, hypothesized drivers of NPP can be assessed via the general equation

$$\text{NPP} = P^{\alpha_P} l_{\text{gs}}^{\alpha_{l_{\text{gs}}}} a^{\alpha_a} e^{-E/kT} g_1 \frac{c_n}{A} \left[\frac{5c_m}{3c_n} \right]^{\alpha} M_{\text{tot}}^{\alpha} \quad (1) \quad \text{and}$$

Here, the dependencies of NPP on precipitation P (mm), growing season length l_{gs} (months (mo) yr^{-1}) and plant age a (yr) are characterized as

power laws with exponents α_P , $\alpha_{l_{\text{gs}}}$ and α_a , respectively (Extended Data Fig. 1). This derivation permits evaluation of nonlinear relationships as well as a hypothesized direct proportionality between NPP and l_{gs} (that is, $\alpha_{l_{\text{gs}}} = 1$)^{7,8}. The influence of temperature T (K) is characterized by an Arrhenius relation with an activation energy E (eV) and Boltzmann's constant k ($8.617 \times 10^{-5} \text{ eV K}^{-1}$). An activation energy of 0.32 eV has been hypothesized for the kinetics of photosynthesis¹⁵. The influences of plant size and stand biomass are described by a size-corrected measure of the stand size distribution c_n (where $f(r) = dn/dr = c_n r^{-\alpha}$ and r is stem radius; m), a normalization constant c_m relating stem radius to plant mass ($r = c_m m^{3/8}$; $\text{m g}^{-3/8}$), the total stand biomass M_{tot} (g), the stand area A (m^2), and a growth normalization constant g_1 ($\text{g m}^{-1-\alpha(5/3)} \text{yr}^{-1+\alpha_{l_{\text{gs}}}-\alpha_a} \text{mm}^{-\alpha_P} \text{mo}^{-\alpha_{l_{\text{gs}}}}$).

To best test metabolic scaling theory predictions and to evaluate direct and indirect climate effects, we recast equation (1) to give a more instantaneous monthly net primary production (NPP/l_{gs} ; $\text{g m}^{-2} \text{mo}^{-1}$) as

$$\frac{\text{NPP}}{l_{\text{gs}}} = P^{\alpha_P} a^{\alpha_a} e^{-E/kT} g_2 \frac{c_n}{A} \left[\frac{5c_m}{3c_n} \right]^{\alpha} M_{\text{tot}}^{\alpha} \quad (2)$$

where l_{gs} (mo yr^{-1}) is growing season length and g_2 is another growth normalization constant ($\text{g m}^{-1-\alpha(5/3)} \text{mo}^{-1-\alpha_{l_{\text{gs}}}} \text{mm}^{-\alpha_P} \text{yr}^{-\alpha_a}$). As discussed below, g_1 and g_2 are governed by several prominent functional and physiological traits^{13,16} and may thus vary with stand characteristics such as soil fertility, leaf type and biome. Equations (1) and (2) can be linearized, respectively, as

$$\ln(\text{NPP}) = \beta_{0,1} + \alpha_P \ln(P) + \alpha_{l_{\text{gs}}} \ln(l_{\text{gs}}) + \alpha_a \ln(a) - \frac{E}{kT} + \alpha \ln(M_{\text{tot}}) \quad (3)$$

$$\ln\left(\frac{\text{NPP}}{l_{\text{gs}}}\right) = \beta_{0,2} + \alpha_P \ln(P) + \alpha_a \ln(a) - \frac{E}{kT} + \alpha \ln(M_{\text{tot}}) \quad (4)$$

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. ²Key Laboratory of Humid Subtropical Eco-geographical Process, Fujian Normal University, Ministry of Education, Fuzhou, Fujian Province 350007, China. ³Department of Biology, Kenyon College, Gambier, Ohio 43022, USA. ⁴The Santa Fe Institute, USA, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA. ⁵The iPlant Collaborative, Thomas W. Keating Bioresearch Building, 1657 East Helen Street, Tucson, Arizona 85721, USA. ⁶Aspen Center for Environmental Studies, 100 Puppy Smith Street, Aspen, Colorado 81611, USA.

where the intercepts are $\beta_{0,1} = \ln(g_1) + \ln(c_n/A) + \ln\left([5c_m^{8/3}/3c_n]^\alpha\right)$ and $\beta_{0,2} = \ln(g_2) + \ln(c_n/A) + \ln\left([5c_m^{8/3}/3c_n]^\alpha\right)$.

Evaluating hypothesized drivers of NPP

Here we use equations (3) and (4) to evaluate several long-standing hypotheses for direct and indirect climate effects on global variation in NPP and NPP/I_{gs} . Specifically, we conducted four separate analyses of globally distributed data on woody plant production. Our data set spans broad ranges of temperature and precipitation (Fig. 1a).

First, in agreement with previous reports^{2–5}, NPP is a significant correlate of mean annual temperature and mean annual precipitation (Fig. 1b, c and Extended Data Table 1). Furthermore, NPP and average annual temperature $<1/kT>_a$ followed an Arrhenius relationship (Fig. 2a and Extended Data Table 1) with an estimated activation energy ($E = 0.296$; 95% confidence interval (CI) = 0.268 to 0.324) that was not different from the hypothesized¹⁵ 0.32 eV. Additionally, NPP was significantly influenced by stand biomass and decreased with plant age (Fig. 1d and Extended Data Table 2).

Second, we examined NPP/I_{gs} (Fig. 2) to assess how climate and ecosystem variables influenced more instantaneous production rates. In contrast to results for NPP, average growing season temperature $<1/kT>_{\text{gs}}$, mean annual precipitation, and mean growing season precipitation explained little to no variation in global NPP/I_{gs} (Fig. 2b–d and Extended Data Table 1). The relationship between NPP/I_{gs} and average growing season temperature (Fig. 2b) provided an estimate of $E = -0.067$ eV (95% CI = -0.131 to 0.003) that was significantly lower than and opposite in direction to the hypothesized 0.32 eV (ref. 15). This suggests that

local adaptation may dampen the ambient temperature response between communities and that the correlation between NPP and mean annual temperature (Fig. 1b) is spurious. For example, mean annual temperature was strongly correlated with growing season length (Extended Data Fig. 2).

Third, to assess hypothesized drivers of NPP while simultaneously controlling for the influence of all other model covariates, we fitted the complete model (equations (1) and (3)) to the data using multiple regression (Table 1 and Extended Data Fig. 3). A large proportion of variation in NPP (full model adjusted $R^2 = 0.769$) was explained by just two variables: stand biomass and plant age. Importantly, in contrast to pairwise correlations (Fig. 1b, c), this multivariate approach revealed that average growing season temperature (partial $r^2 = 0.073$) and mean growing season precipitation (partial $r^2 = 0.011$) explained little of the variation in NPP, and growing season length explained almost none (partial $r^2 = 0.004$). The mass-scaling exponent was estimated as $\alpha = 0.763$ (95% CI = 0.735 to 0.792), significantly greater than the metabolic scaling theory prediction of $3/5 = 0.60$. Furthermore, the scaling exponent for growing season length ($\alpha_{\text{gs}} = 0.058$; 95% CI = -0.007 to 0.109) was significantly lower than the value of 1 required for direct proportionality^{7,8}. The estimated activation energy $E = 0.195$ eV (95% CI = 0.156 to 0.234) did not include the hypothesized 0.32 eV. These general conclusions did not change when using mean annual temperature and/or mean annual precipitation, or root, aboveground woody, and foliage components of NPP (Extended Data Table 3).

Comparing the fit of the complete model (equations (1) and (3)) with a simpler model containing only plant age and stand biomass ($\text{NPP} = ca^{\alpha_a} M_{\text{tot}}^\alpha$) yielded similar coefficients of determination but a higher Akaike information criterion (AIC) value for the complete model (AIC = 140.115,

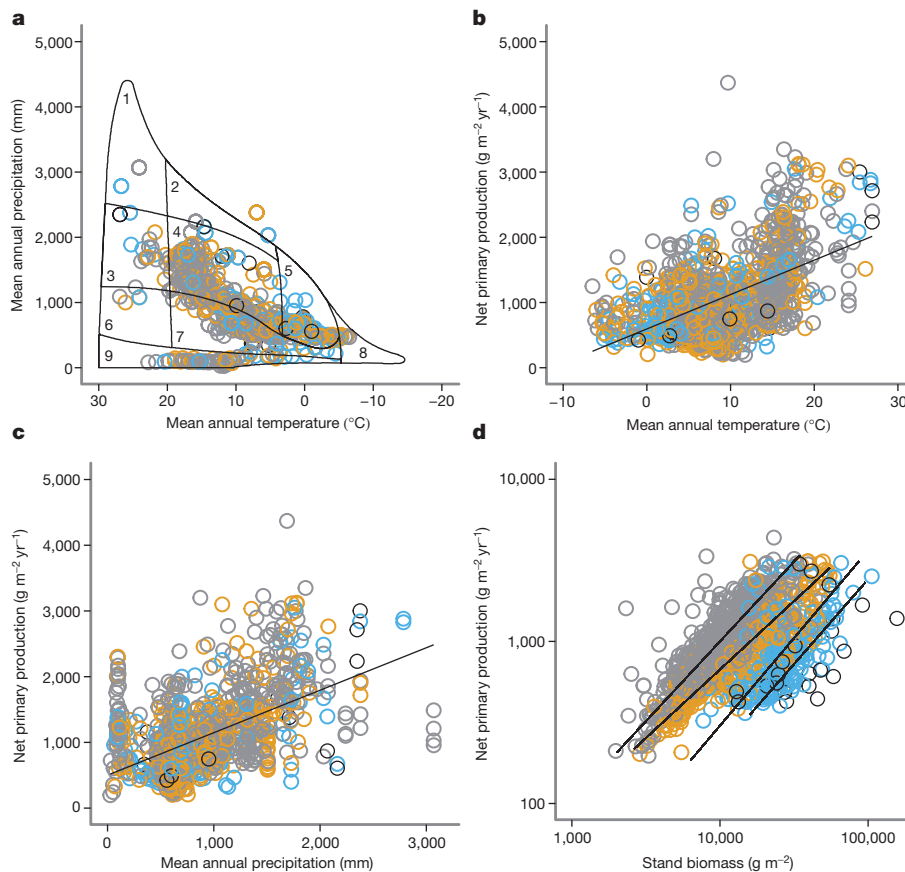


Figure 1 | Global variation in annual net primary production for 1,247 woody plant communities grouped by age class. **a**, Precipitation–temperature space occupied by the plant communities. Biome definitions from ref. 39. 1, tropical rainforest; 2, temperate rainforest; 3, tropical seasonal forest; 4, temperate forest; 5, taiga; 6, savannah; 7, woodland/shrubland; 8,

tundra; 9, desert. **b**, Relationship between NPP and mean annual temperature. **c**, Relationship between NPP and mean annual precipitation. **d**, Relationship between NPP and stand biomass. Grey, 0–50 years; orange, 51–100 years; blue, 101–200 years; black, ≥201 years.

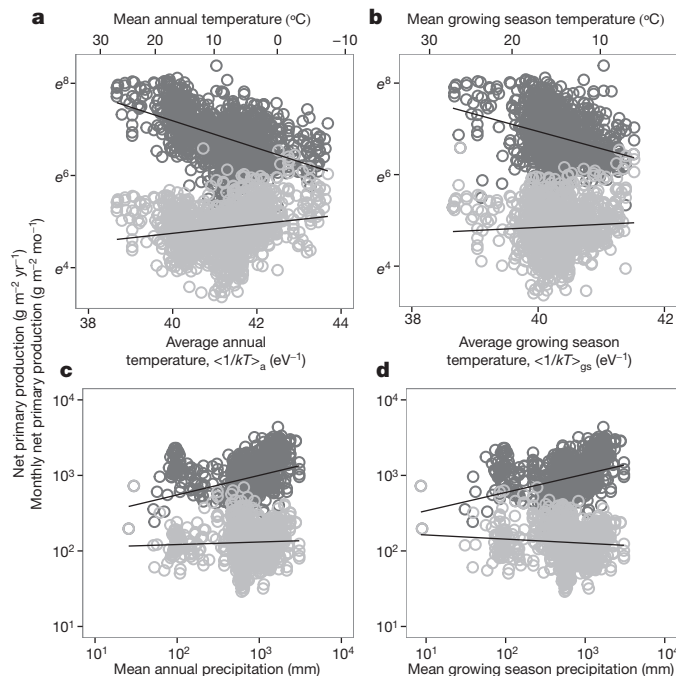


Figure 2 | Net primary production of woody plant communities across global climate gradients. **a**, Annual temperature. **b**, Growing season temperature. **c**, Annual precipitation. **d**, Growing season precipitation. Dark grey, annual net primary production (NPP); light grey, monthly net primary production (NPP/ l_{gs}); e , mathematical constant (~ 2.718).

$R^2 = 0.769$) than for the simpler model (AIC = -1768.585 , $R^2 = 0.735$; Fig. 3a and Extended Data Fig. 4), indicating that age and biomass together explained most of the variation in NPP (Fig. 3a). Soil and leaf trait differences did not influence scaling relationships, but did influence normalization constants (Fig. 3b, c and Supplementary Information). Biome differences affected scaling relationships in some cases and normalization constants in others (Extended Data Fig. 4 and Supplementary Information).

Fourth, because metabolic scaling theory predictions are based on instantaneous rates, a more precise evaluation should consider instantaneous NPP with climate variables most relevant to growth physiology. Therefore, we used multiple regression to assess metabolic scaling theory for rates of NPP/l_{gs} (equations (2) and (4)) using average growing season temperature and mean growing season precipitation (Table 1 and Fig. 4). With these data, all covariates were significant (Table 1). The explanatory ability of this model was reduced (adjusted $r^2 = 0.440$), probably due to error in growing season length estimates. Similar to NPP, stand biomass and plant age were the best predictors of variation in NPP/l_{gs} , with little of the variation explained by mean growing season precipitation (partial $r^2 = 0.095$) and essentially none of the variation explained by average growing season temperature (partial $r^2 = 0.007$). In support of metabolic scaling theory, the mass-scaling exponent was estimated as $\alpha = 0.613$ (95% CI = 0.575 to 0.652), which is indistinguishable from the metabolic scaling theory prediction of $3/5 = 0.60$. However, the estimated activation energy $E = -0.079$ (95% CI = -0.130 to 0.028) was significantly lower and opposite in direction to the hypothesized value of $E = 0.32$ eV (ref. 15). These conclusions did not generally change when using root, aboveground woody and foliage components of NPP/l_{gs} (Extended Data Table 3).

Climate has little direct effect on NPP

Our analyses question several long-held and more recent conclusions regarding the influence of climate on global variation in NPP, and reveal that a number of central conclusions established in studies using bivariate regression are probably spurious. For example, mean annual temperature and mean annual precipitation are often cited as primary drivers of

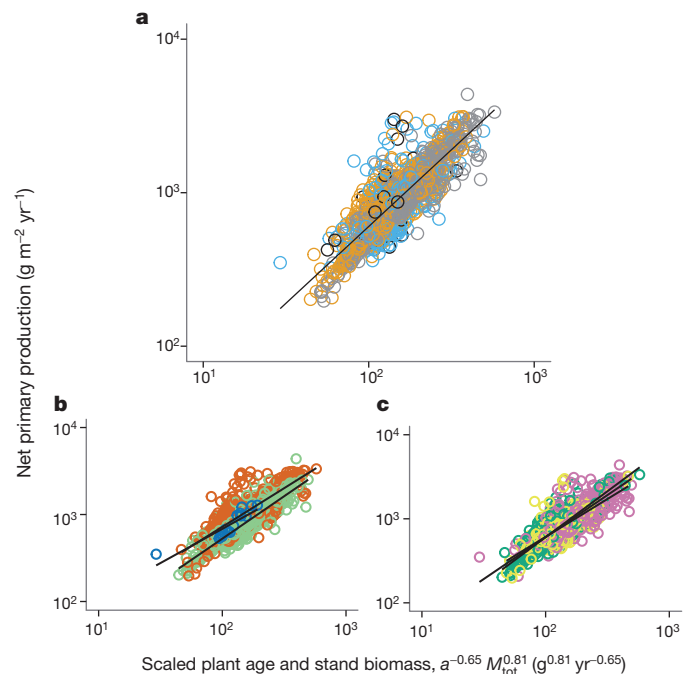


Figure 3 | Global variation in annual net primary production of woody plant communities expressed as a general scaling function of plant age and stand biomass M_{tot} . **a**, 1,247 stands grouped by age class with ordinary least squares (OLS) regression line; **b**, 1,247 stands grouped by leaf functional trait type with standardized major axis (SMA) regression lines; **c**, 1,237 stands grouped by soil fertility class with SMA regression lines. Grey, 0–50 years; light orange, 51–100 years; light blue, 101–200 years; black, ≥ 201 years; dark orange, broadleaf; light green, needle-leaf; dark blue, mixed-leaf; pink, low soil fertility; yellow, medium soil fertility; dark green, high soil fertility.

NPP (see refs 2–5 and Fig. 1b, c), but after controlling for plant age and stand biomass, temperature and precipitation explained little to none of the variation (Table 1 and Fig. 4a, b). Likewise, hypothesized activation values of 0.32 eV have been supported by bivariate relationships (for example, Fig. 2a and ref. 15), but accounting for other covariates yielded estimates that did not support these predictions¹⁵ (Table 1 and Fig. 4a). These results are intriguing given that the temperature-dependencies of local scale photosynthesis and respiration are well established⁶.

Three factors might account for the absence of direct climate effects. First, like most studies in plant ecology and ecosystem metabolism^{2,3,7,10,11,16}, our analyses considered ambient air temperature. However, air temperatures may not reflect the tissue temperatures that govern plant growth rates, because plant traits (thermophysical properties) can influence energy budgets and decouple plant tissue temperatures from air temperature¹⁷. This could dampen ambient temperature correlations across climate gradients if, for example, selection adjusts leaf traits to maintain leaf temperatures near photosynthetic optima¹⁸. As air temperature is one of the most commonly used climate variables in plant and ecosystem ecology, any directional plant–air temperature differences (as recently suggested¹⁸) will have profound implications for our understanding of plant–climate interactions. Mean annual air temperature may yield especially misleading results, because it can differ substantially from the operative temperatures of organisms, and it also covaries with other key drivers of metabolism and NPP (see, for example, Extended Data Fig. 2) that can act as confounding variables to produce spurious relationships. We suggest that future studies move away from using mean annual air temperature and instead use air and plant body temperatures measured during the growing season and/or key periods of development. Second, biochemical adaptation and/or acclimatization to cold temperatures may increase plant metabolism^{7,8,19}. For example, observed shifts in foliar chemistry and metabolic efficiencies have been argued to offset variation in metabolic kinetics across temperature gradients^{8,13}. Third, the climate

Table 1 | Multiple regression fits of theory (equations (3) and (4)) to a global compilation of data from 1,247 woody plant communities

Dependent variable	Covariate	Coefficient	Estimate	95% CI	s.e.	t	P value	Partial r^2
NPP ($\text{g m}^{-2} \text{yr}^{-1}$)	$<1/kT>_{\text{gs}}$	β_0	9.336	7.758 to 10.914	0.804	11.609	$<2 \times 10^{-16}$	0.098
		E	0.195	0.156 to 0.234	0.020	9.854	$<2 \times 10^{-16}$	0.073
	a	α_a	-0.568	-0.599 to -0.537	0.016	-35.808	$<2 \times 10^{-16}$	0.508
	l_{gs}	$\alpha_{l_{\text{gs}}}$	0.058	0.007 to 0.109	0.026	2.223	0.026	0.004
	M_{tot}	α	0.763	0.735 to 0.792	0.014	52.863	$<2 \times 10^{-16}$	0.693
	P_{gs}	α_P	0.043	0.020 to 0.067	0.012	3.664	2.58×10^{-4}	0.011
NPP/ l_{gs} ($\text{g m}^{-2} \text{mo}^{-1}$)	$<1/kT>_{\text{gs}}$	β_0	-1.652	-3.741 to 0.438	1.065	-1.551	0.121	0.002
		E	-0.079	-0.130 to 0.028	0.026	3.016	0.003	0.007
	a	α_a	-0.168	-0.392 to -0.310	0.021	-16.711	$<2 \times 10^{-16}$	0.184
	M_{tot}	α	0.613	0.575 to 0.652	0.020	30.995	$<2 \times 10^{-16}$	0.436
	P_{gs}	α_P	-0.168	-0.197 to -0.139	0.015	-11.444	$<2 \times 10^{-16}$	0.095

$<1/kT>_{\text{gs}}$, average growing season temperature; a , age; E , activation energy; l_{gs} , growing season length; M_{tot} , stand biomass; P_{gs} , mean growing season precipitation; α , mass scaling exponent; α_a , age scaling exponent; $\alpha_{l_{\text{gs}}}$, growing season length scaling exponent; α_P , precipitation scaling exponent; β_0 , intercept.

data used here were interpolated from 29-year climate station means²⁰ and are not necessarily representative of the years when production data were obtained. Thus, regressing short-term NPP estimates on longer-term climate estimates will add noise to the relationships. Nonetheless, while we are unaware if this error yields a directional bias across climate gradients, future work should assess its importance.

Terrestrial NPP increases asymptotically with precipitation^{4,5} because plant growth in terrestrial ecosystems is generally water limited. However, our analyses suggest that this relationship doesn't occur through direct effects of precipitation on plant metabolism per se, but instead via indirect effects of water availability on stand biomass and plant age. Although this is counterintuitive given the importance of water in whole-plant physiology (for example, avoidance of xylem embolism and control of stomatal aperture), it is consistent with a more hydrological view of plant-atmosphere interactions²¹. Specifically, biomass controls the total leaf area that drives transpiration, assimilation and growth. Furthermore, precipitation is not necessarily representative of plant-available water²², so rather than using precipitation as the primary measure of plant-available water, evapotranspiration⁵ should also be included.

Plant age effects on NPP

Even after controlling for stand biomass and climate, global NPP/ l_{gs} decreased with age so that (for a given biomass) younger stands had

higher rates of production. Although such age-related declines are well documented²³, the drivers of this pattern remain unclear. Numerous mechanisms have been proposed, including: (1) hydraulic limitation of plant height²⁴; (2) changes in carbon use efficiency²⁵, potentially from increasing respiration requirements with size; (3) ontogenetic shifts in biomass allocation, with smaller plants having proportionately more foliage and higher assimilation rates; (4) increasing light limitation with age^{26,27}; and (5) decreasing stand density with age²⁸, which would act through the size-corrected size distribution term c_n . Together, our results reiterate a need in global change studies for a deeper understanding of the causal mechanisms linking age to plant production.

Climate controls NPP via biomass and age

Whereas temperature and water availability are fundamental drivers of plant physiology and ecosystem metabolism at local scales^{19,29,30}, at global scales they appear to have little direct kinetic control on NPP. Instead, our findings suggest that climate influences NPP indirectly via plant age and stand biomass (see ref. 11), which is largely driven by maximum plant size^{31,32}. For example, plant age is influenced by time since last disturbance, and maximum plant size is constrained by limitations on the water and energy fluxes necessary to support basal metabolism^{24,31,32}.

Our theoretical framework further extends the predictive ability of metabolic scaling theory and links multiple hypothesized climate drivers to plant and ecosystem metabolism. Furthermore, it uniquely (1) integrates metabolic scaling and physiological approaches to plant ecology and global change studies; (2) underscores the importance of plant size, allometry and age as primary drivers of variation in ecosystem metabolism; and (3) provides a foundation to assess if adaptive differences in plant form and function can compensate for broad-scale climate gradients. Although our results support the mechanistic basis of metabolic scaling theory for the origin of the ecosystem mass-scaling exponent α (see ref. 11) and the influence of plant traits on variation in the normalization constants g_1 and g_2 , they also highlight a need for integrative theory that explains the age-dependence of terrestrial NPP (estimated here as $\alpha_a = -0.65$). Additionally, future work is needed on the geographical scale at which the activation energy for plant metabolism becomes decoupled from temperature.

We have shown that global variation in terrestrial NPP is consistent with the hypothesis that the diversity of plant form and function originated via selection to maximize plant growth across climate gradients^{7,8} (Fig. 4). This has resulted in convergence to a common scaling relationship between NPP, plant age and total stand biomass (Fig. 3). Interestingly, recent analyses indicate that both mean annual temperature and mean annual precipitation are also poor predictors of total stand biomass³² (but see ref. 33). Additionally, metabolic scaling theory predicts^{11,12} and recent empirical data show^{32,34} that the best predictor of total autotrophic ecosystem biomass appears to be the size of the largest individual. Consequently, efforts to predict ecosystem function in response to global change should include the mechanisms that govern maximum plant size³¹ and general ecosystem scaling relationships (Fig. 3).

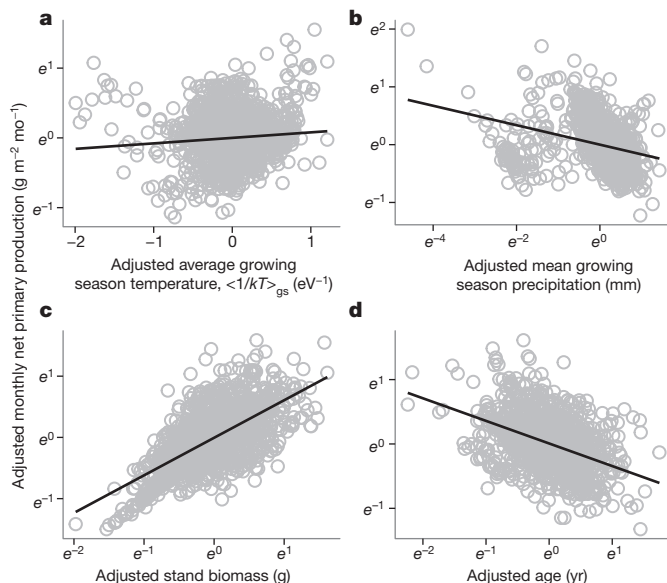


Figure 4 | Partial regression plots illustrating relationships between monthly net primary production (NPP/ l_{gs}) and individual covariates from equation (4) for 1,247 woody plant communities. Plots show the correct relationship (slope and variance) between NPP/ l_{gs} and each covariate while controlling for the influence of all other model covariates. **a**, Average growing season temperature; **b**, mean growing season precipitation; **c**, stand biomass; **d**, plant age; e , mathematical constant (~ 2.718).

METHODS SUMMARY

We assessed variation in NPP and NPP/I_{gs} across broad climate gradients using a global compilation of biomass and production data for 1,247 woody plant communities^{8,35–38} (and Malhi, Y. *et al.*, submitted) and climate data from a high-resolution gridded data set²⁰ (see Methods). NPP was computed as the sum of annual production of root, stem, branch, reproductive (when available) and foliage components. To calculate NPP/I_{gs} , growing season length was calculated as the number of months with a mean minimum temperature greater than 0.6 °C and a moisture index $\text{MI} > 0.048$ (ref. 8). Temperature and precipitation were calculated as both annual and growing season averages, and temperature was also expressed as the Boltzmann factor exponent $1/kT$. Relationships between production and climate variables were first assessed using OLS linear regression. Next, equations (3) and (4) were fit to compiled data to evaluate relationships between NPP or NPP/I_{gs} and hypothesized drivers. Finally, the fit of the complete model (equations (1) and (3)) was compared with that of a simpler model ($\text{NPP} = ca^{2a} M_{\text{tot}}^Z$) obtained via multiple regression with age and biomass as covariates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2013; accepted 12 May 2014.

Published online 20 July; corrected online 6 August 2014 (see full-text HTML version for details).

- Schimel, D. S. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**, 169–172 (2001).
- Lieth, H. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) (Springer, 1975).
- Schuur, E. A. G. Productivity and global climate revisited: The sensitivity of tropical forest growth to precipitation. *Ecology* **84**, 1165–1170 (2003).
- Huxman, T. E. *et al.* Convergence across biomes to a common rain-use efficiency. *Nature* **429**, 651–654 (2004).
- Ponce Campos, G. E. *et al.* Ecosystem resilience despite large-scale altered hydroclimatic conditions. *Nature* **494**, 349–352 (2013).
- Berry, J. & Björkman, O. Photosynthetic response and adaptation to temperature in higher plants. *Annu. Rev. Plant Physiol.* **31**, 491–543 (1980).
- Enquist, B. J., Kerkhoff, A. J., Huxman, T. E. & Ecosystem, E. P. Adaptive differences in plant physiology and ecosystem paradoxes: insights from metabolic scaling theory. *Glob. Change Biol.* **13**, 591–609 (2007).
- Kerkhoff, A. J., Enquist, B. J., Elser, J. J. & Fagan, W. F. Plant allometry, stoichiometry and the temperature-dependence of primary productivity. *Glob. Ecol. Biogeogr.* **14**, 585–598 (2005).
- Bonan, G. B. Physiological derivation of the observed relationship between net primary production and mean annual air temperature. *Tellus B Chem. Phys. Meteorol.* **45**, 397–408 (1993).
- Chapin, F. S. Effects of plant traits on ecosystem and regional processes: A conceptual framework for predicting the consequences of global change. *Ann. Bot. (Lond.)* **91**, 455–463 (2003).
- Enquist, B. J., West, G. B. & Brown, J. H. Extensions and evaluations of a general quantitative theory of forest structure and dynamics. *Proc. Natl Acad. Sci. USA* **106**, 7046–7051 (2009).
- West, G. B., Enquist, B. J. & Brown, J. H. A general quantitative theory of forest structure and dynamics. *Proc. Natl Acad. Sci. USA* **106**, 7040–7045 (2009).
- Enquist, B. J. *et al.* A general integrative model for scaling plant growth, carbon flux, and functional trait spectra. *Nature* **449**, 218–222 (2007).
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M. & West, G. B. Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).
- Allen, A. P., Gillooly, J. F. & Brown, J. H. Linking the global carbon cycle to individual metabolism. *Funct. Ecol.* **19**, 202–213 (2005).
- Enquist, B. J. & Bentley, L. P. in *Metabolic Ecology: A Scaling Approach* (eds Sibily, R. M., Brown, S. & Kodric-Brown, A.) 164–187 (Wiley, 2012).
- Gates, D. M. *Biophysical Ecology* (Springer, 1980).
- Helliker, B. R. & Richter, S. L. Subtropical to boreal convergence of tree-leaf temperatures. *Nature* **454**, 511–514 (2008).
- Atkin, O. K. & Tjoelker, M. G. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends Plant Sci.* **8**, 343–351 (2003).
- New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25 (2002).
- Monteith, J. L. Climate and the efficiency of crop production in Britain. *Phil. Trans. R. Soc. Lond. B* **281**, 277–294 (1977).
- Renee Brooks, J., Barnard, H. R., Coulombe, R. & McDonnell, J. J. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geosci.* **3**, 100–104 (2010).
- Gower, S. T., McMurtrie, R. E. & Murty, D. Aboveground net primary production decline with stand age: potential causes. *Trends Ecol. Evol.* **11**, 378–382 (1996).
- Ryan, M. G. & Yoder, B. J. Hydraulic limits to tree height and tree growth. *Bioscience* **47**, 235–242 (1997).
- DeLucia, E. H., Drake, J. E., Thomas, R. B. & Gonzalez-Meler, M. Forest carbon use efficiency: is respiration a constant fraction of gross primary production? *Glob. Change Biol.* **13**, 1157–1167 (2007).
- Kerkhoff, A. J. & Enquist, B. J. Ecosystem allometry: the scaling of nutrient stocks and primary productivity across plant communities. *Ecol. Lett.* **9**, 419–427 (2006).
- Muller-Landau, H. C. *et al.* Testing metabolic ecology theory for allometric scaling of tree size, growth, and mortality in tropical forests. *Ecol. Lett.* **9**, 575–588 (2006).
- Niklas, K. J., Midgley, J. J. & Rand, R. H. Tree size frequency distributions, plant density, age and community disturbance. *Ecol. Lett.* **6**, 405–411 (2003).
- Yvon-Durocher, G. *et al.* Reconciling the temperature dependence of respiration across timescales and ecosystem types. *Nature* **487**, 472–476 (2012).
- Wright, I. J. & Westoby, M. Differences in seedling growth behaviour among species: trait correlations across species, and trait shifts along nutrient compared to rainfall gradients. *J. Ecol.* **87**, 85–97 (1999).
- Kempes, C. P., West, G. B., Crowell, K. & Girvan, M. Predicting maximum tree heights and other traits from allometric scaling and resource limitations. *PLoS ONE* **6**, e20551 (2011).
- Stegen, J. C. *et al.* Variation in above-ground forest biomass across broad climatic gradients. *Glob. Ecol. Biogeogr.* **20**, 744–754 (2011).
- Larjavaara, M. & Muller-Landau, H. C. Rethinking the value of high wood density. *Funct. Ecol.* **24**, 701–705 (2010).
- Stephenson, N. L. *et al.* Rate of tree carbon accumulation increases continuously with tree size. *Nature* **507**, 90–93 (2014).
- Cannell, M. G. R. *World Forest Biomass and Primary Production Data* (Academic, 1982).
- Luo, T. X. *Patterns of Biological Production and its Mathematical Models for Main Forest Types of China*. PhD thesis, Chinese Academy of Sciences (1996).
- Clark, D. S. *et al.* Net primary productivity in tropical forests: An evaluation and synthesis of existing field data. *Ecol. Appl.* **11**, 371–384 (2001).
- Luyssaert, S. *et al.* CO₂ balance of boreal, temperate, and tropical forests derived from a global database. *Glob. Change Biol.* **13**, 2509–2537 (2007).
- Whittaker, R. H. *Communities and Ecosystems* (Macmillan, 1970).

Supplementary Information is available in the online version of the paper.

Acknowledgements S.T.M. and B.J.E. were supported by an NSF MacroSystems award (1065861) and a fellowship from the Aspen Center for Environmental Studies. D.C. was supported by the National Natural Science Foundation of China (31170374 and 31370589) and Fujian Natural Science Funds for Distinguished Young Scholar (2013J06009). A.J.K. was supported by a sabbatical supplement from Kenyon College, and by a National Science Foundation ROA supplement (1065861) to the NSF MacroSystems award (1065861) to B.J.E.

Author Contributions S.T.M., D.C., A.J.K. and B.J.E. compiled data, developed theory, performed analyses and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.J.E. (benquist@email.arizona.edu) or D.C. (chengdl@fjnu.edu.cn).

Replaying evolutionary transitions from the dental fossil record

Enni Harjunmaa¹, Kerstin Seidel^{2,3}, Teemu Häkkinen¹, Elodie Renvoisé¹, Ian J. Corfe¹, Aki Kallonen⁴, Zhao-Qun Zhang⁵, Alistair R. Evans^{6,7}, Marja L. Mikkola¹, Isaac Salazar-Ciudad^{1,8}, Ophir D. Klein^{2,3,9,10} & Jukka Jernvall¹

The evolutionary relationships of extinct species are ascertained primarily through the analysis of morphological characters. Character inter-dependencies can have a substantial effect on evolutionary interpretations, but the developmental underpinnings of character inter-dependence remain obscure because experiments frequently do not provide detailed resolution of morphological characters. Here we show experimentally and computationally how gradual modification of development differentially affects characters in the mouse dentition. We found that intermediate phenotypes could be produced by gradually adding ectodysplasin A (EDA) protein in culture to tooth explants carrying a null mutation in the tooth-patterning gene *Eda*. By identifying development-based character inter-dependencies, we show how to predict morphological patterns of teeth among mammalian species. Finally, *in vivo* inhibition of sonic hedgehog signalling in *Eda* null teeth enabled us to reproduce characters deep in the rodent ancestry. Taken together, evolutionarily informative transitions can be experimentally reproduced, thereby providing development-based expectations for character-state transitions used in evolutionary studies.

In the case of extinct mammals, a large number of dental features are used as characters in phylogenetic analyses^{1–4}, and these characters often provide the key evidence for evolutionary inferences due to the preponderance of teeth in the fossil record. For reliable phylogenetic inferences, characters have been typically considered to be independent from each other^{5–8}. Although developmental factors can make characters dependent^{9–11}, thorough analyses of the influence of development on character state changes are lacking. To approximate changes relevant to evolutionary transitions, experiments that tune morphology gradually are required. These kinds of experiments are also useful to evaluate how, and whether, continuous changes in underlying developmental or genetic parameters map to continuous changes in the phenotype^{12–14}.

Here we investigated whether gradual alterations of tooth development can produce gradual changes in the phenotype, and whether these changes reflect known evolutionary transitions. We focused on the development of the rodent dentition, using mice carrying a spontaneously occurring null mutation in ectodysplasin (*Eda*) as a starting point. This mutation was chosen because the effects of *Eda* on tooth morphology are relatively subtle, causing simplification of dental morphology without complete loss of teeth^{10,15}, however, the *Eda* mutation causes changes in many characters and is thus highly informative¹⁰.

Experimental tuning of morphology

We reasoned that, to approximate evolutionary transitions, fine-tuning of EDA signalling would be required. We tracked gradual changes during development by crossing *Eda* null mice with mice that express green fluorescent protein (GFP) from the *Shh* locus (hereafter called ShhGFP mice¹⁶). The epifluorescence of ShhGFP mice can be used to monitor tooth cusp development because *Shh* is initially expressed in the enamel knots, which are the epithelial signalling centres that form at the positions of future cusps¹⁷. Later during differentiation, *Shh* expression spreads

throughout the inner enamel epithelium, enabling the visualization of the overall crown shape.

First, we used EDA protein in culture at increasing concentrations ($n = 9$ to 16 in each group, Supplementary Table 1, Methods) to test whether the *Eda* null morphology could be engineered to gradually resemble wild-type morphology. We cultured first lower molars starting at embryonic day 13, just before crown formation begins, and EDA protein was administered into the culture media at days zero and two. This treatment scheme restored EDA signalling during the period of first molar cusp patterning. At this stage, *Eda* is thought to regulate the size and signalling of enamel knots¹⁰, which in turn give rise to tooth cusps.

The EDA protein treatments restored the wild-type mouse cusp pattern in culture (Fig. 1a), in agreement with previous experiments^{18,19}. We next examined the mode of cusp appearance in detail by analysing daily time-lapse images of the cultured teeth. The results showed that increasing dosage of EDA caused a heterochronic shift in cusp initiation (Fig. 1a). Specifically, some of the cusps were initiated earlier (predisplaced) as EDA concentration was increased (Fig. 1a, Supplementary Table 1). Furthermore, the time-lapse data showed that increasing EDA concentration enlarged the primary enamel knot, which in turn increased the number of cusps (Fig. 1b, Extended Data Fig. 1). The link between the primary enamel knot size and cusp number (Fig. 1b) indicates that the overall size of the tooth crown needs to reach certain thresholds to accommodate additional cusps. From a developmental signalling point of view, a heterometric²⁰ change in the dosage of EDA signalling can lead to a heterochronic shift in timing of cusp initiation.

Computational modelling of patterning

Computational modelling has been used to simulate tooth shape development and evolution^{14,21}, and the new developmental data that we obtained allowed us to link models and experiments in unprecedented detail (see

¹Developmental Biology Program, Institute of Biotechnology, University of Helsinki, P.O. Box 56, FIN-00014 Helsinki, Finland. ²Program in Craniofacial and Mesenchymal Biology, University of California, San Francisco, San Francisco, California 94114, USA. ³Department of Orofacial Sciences, University of California, San Francisco, San Francisco, California 94114, USA. ⁴Division of Materials Physics, Department of Physics, University of Helsinki, P.O. Box 64, FIN-00014 Helsinki, Finland. ⁵Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. ⁶School of Biological Sciences, Monash University, Victoria 3800, Australia. ⁷Geosciences, Museum Victoria, GPO Box 666, Melbourne, Victoria 3001, Australia. ⁸Genomics, Bioinformatics and Evolution Group, Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès 08193, Spain. ⁹Department of Pediatrics, University of California, San Francisco, San Francisco, California 94114, USA. ¹⁰Institute for Human Genetics, University of California, San Francisco, San Francisco, California 94114, USA.

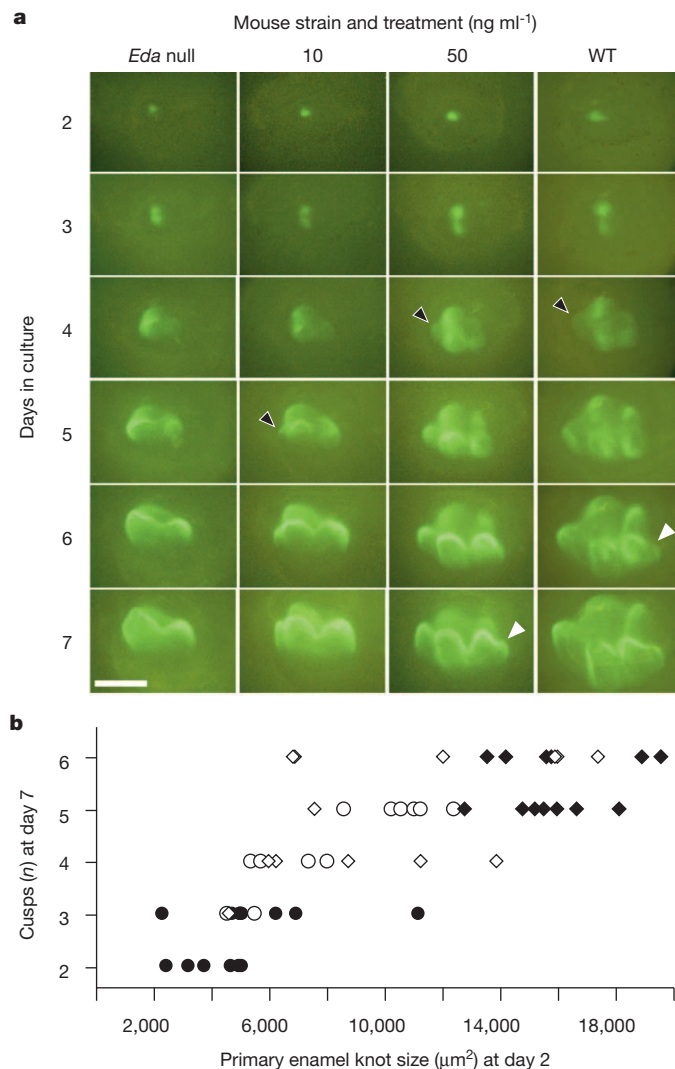


Figure 1 | Gradual dosage effects of EDA on *Eda* null mutant first lower molars (m1). **a**, *ShhGFP* × *Eda* null tooth development is rescued by EDA, with higher concentrations reproducing wild type (WT) development (*Eda* null: $n = 15$; 10 ng ml⁻¹: $n = 15$; 50 ng ml⁻¹: $n = 13$; WT: $n = 16$; all teeth listed in Supplementary Table 1: $n = 113$). Initiation of different parts of the tooth crown is shifted earlier with higher EDA concentrations, such as the anteroconid (black arrowheads) and the hypoconulid (white arrowheads). **b**, Primary enamel knot size at culture day 2 predicts the number of cusps at day 7. *Eda* null teeth treated with 10 ng ml⁻¹ (open circles) and 50 ng ml⁻¹ (open diamonds) fill in the phenotypic gap between the *Eda* null (black circles) and wild-type (black diamonds) teeth. Anterior is towards the left in **a**. Scale bar, 500 μm.

Methods). To model the experimental transitions, we implemented a morphodynamic model, which integrates signalling and tissue growth to simulate tooth development²¹, in the new ToothMaker interface (Extended Data Fig. 2). First we modelled a wild-type mouse tooth morphology corresponding to our cultured teeth (see Methods). Then, by progressively adjusting (mutating) each parameter separately, we simulated the effects of gradual changes in signalling (Extended Data Fig. 3). The *in silico* simulation reproduced the fusion of cusps both on the talonid and on the trigonid of the *Eda* null model, as well as rescue of the separate cusps observed in the *in vitro* experiment (Fig. 2a, Extended Data Fig. 4). Moreover, as predicted from the experimental observations, the full range of transitions from fused trigonid to separate cusps was replicated by varying the activator parameter that induces the formation of enamel knots (Fig. 1a, Extended Data Fig. 4).

The comparable patterns in the experiments (Fig. 1) and the modelling (Fig. 2) underscore the dynamic nature of tooth shape development,

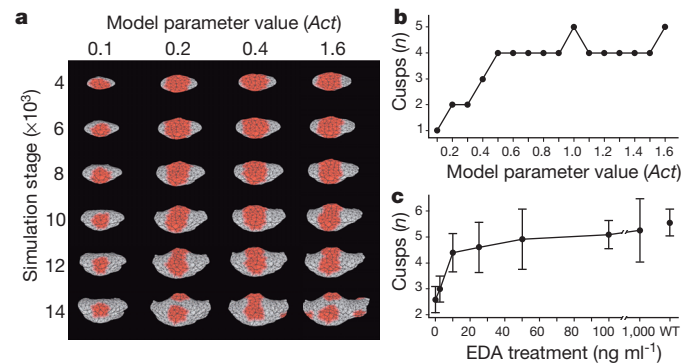


Figure 2 | Computational modelling of gradual changes in signalling on cusp patterns. **a**, Computer simulations using ToothMaker (Extended Data Fig. 2) of first lower molar development show appearance of cusp areas (red colour). Larger values of activator (*Act*) parameter increase cusps. **b**, Modelled wild-type mouse pattern is largely retained until $Act = 0.5$. Smaller *Act* values quickly reduce cusps resulting in modelled teeth that resemble *Eda* null teeth. **c**, Tabulated data from culture experiments (Supplementary Table 1) show an abrupt change in cusps at low levels of EDA protein (2.5 to 10 ng ml⁻¹), similar to modelling data (**b**). Increasing EDA concentration does not increase cusps beyond wild type (WT). Error bars denote s.d. Anterior is towards the left in **a**.

and indicate that the identification of cusp homologies should rely on topological correspondences rather than unique, cusp specific gene expression patterns. Furthermore, both the model simulations and the experiments showed a rapid phenotypic response at low levels of activator (Fig. 2b) and EDA signalling (Fig. 2c), respectively, with smaller changes observed at higher levels of signalling. These results imply a potential disjunction between rates of evolution measured in the phenotype and in gene expression level, although by varying more than one parameter at a time, a multivariate linear relationship between expression levels and phenotypes remains possible^{12,13,19}.

Detailed analyses of character states

To examine the full range of morphologies produced in the cultured explants, we tabulated the character states comparable to the ones used in evolutionary studies for each crown region (see Methods and Supplementary Table 1). Although the tooth culturing system does not produce mineralized features, we were able to tabulate the presence or absence of cusps (Fig. 3a) and relative height of the talonid (Fig. 3b) with high resolution.

First, the trigonid cusps, the protoconid and the metaconid, are frequently fused in *Eda* null teeth (40% of explants, Fig. 3c, Supplementary Table 1). The separation of the protoconid and the metaconid in 2.5 ng ml⁻¹ EDA treated teeth reflects a transition that would predate the evolution of the pretribosphenic mammalian pattern^{11,21}.

Next, the talonid, which in *Eda* null teeth is a shallow shelf lacking well-defined cusps, was already affected in the lowest, 2.5 ng ml⁻¹ EDA treatments by an increase in height, and in the second lowest 10 ng ml⁻¹ treatments by acquisition of additional cusps (Fig. 3c, Supplementary Table 1). These treatments, however, caused polymorphic effects (Fig. 3c). In 47% of 10 ng ml⁻¹ explants, the shallow shelf gave rise to a single cusp, whereas in the remaining explants two distinct cusps formed. These two talonid cusps correspond to the hypoconid and the entoconid cusps in wild-type teeth, and both cusps formed consistently starting at 100 ng ml⁻¹ EDA (Fig. 3c).

During the evolution of tribospheny in Mesozoic mammals, the functional, three cusped talonid was added to the posterior end of the trigonid^{11,22}. This originated in Triassic non-mammalian synapsids where a single cusp, often located on the cingulid, was appended posteriorly to the basal three-cusped trigonid morphology^{22,23}. Although the hypoconid is generally agreed to have evolved before the entoconid²², our data do not allow determination of whether the single cusp in the talonid of

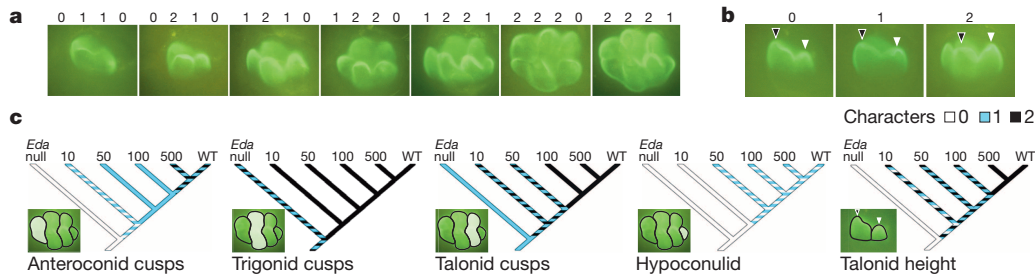


Figure 3 | Differential sensitivities of tooth crown regions to EDA. **a**, Range of tooth morphologies and cuspal character states tabulated at the end of the cultures (day 13 + 7). Character state numbers above first lower molar images correspond to the number of cusps present in the respective region of the crown (Supplementary Table 1). The first trait is for the anteroconid, the second for the trigonid, the third for the talonid and the fourth for the hypoconulid.

cultured teeth is the hypoconid or the entoconid. The central location of the single talonid cusp in our cultured teeth (Figs 1a and 3a) does, however, suggest that the effects of EDA, at least partially, mimic the early steps of talonid evolution.

Finally, the anteroconid and the hypoconulid appear already at 10–50 ng ml⁻¹ EDA, but unlike *in vivo*, their wild-type character states remain polymorphic in cultured molars irrespective of treatments (shown by hatched colouring in Fig. 3c). Evolutionarily, these crown features appear in early rodents and are present in the basal murines^{2,3,24,25}. The hypoconulid, however, has been lost in many murine lineages, and this evolutionary lability appears to be reflected in the developmental data.

Taken together, these data show that even though dental characters used in evolutionary analyses may be highly pleiotropic, as shown previously¹⁰, transformations of character states can occur at different thresholds of signalling (Fig. 3c). In contrast to the relatively robust trigonid cusps, large variation in talonid structure can be reproduced by small changes of EDA signalling. This has major implications for the evolution of tribosphenic mammals that are diagnosed by their derived talonid features^{1,11,22}, as discussed below.

Testing the experimental predictions

To test the development-based predictions of evolutionary patterns in the talonid, we examined the link between talonid height and cusp number across mammalian species. We first tabulated relative talonid height (as percentage of the trigonid height) and cusp numbers in the entire posterior region of the tooth (the entoconid, the hypoconid and the hypoconulid; see Methods). The results show that talonid height and cusp number are linked developmentally (Fig. 4a, Extended Data Table 1), even though these traits are typically treated as independent characters in evolutionary analyses. In the macroevolutionary context of rodents, the patterns obtained in the experiments appear to bridge the derived state we analysed in 35 species of extant murine rodents, already present in the extinct Miocene early murines *Potwarmus* and *Antemus*^{24,25}, and the basal morphology found in *Tribosphenomys minutus*, a Paleocene mammal that is considered to be a basal rodentiaform², or the immediate sister taxon of Glires⁴ (Fig. 4a, Extended Data Table 1).

Because our experimental morphologies extend even beyond those found in rodents and towards further reduced talonids (Fig. 4a), we also measured talonid height and cusp number in 32 species of extant carnivorans. The first lower molar, or carnassial, of carnivorans shows arguably the fullest range of talonid morphologies among extant mammals²⁶. The correlated change in carnivoran talonid height and cusp number (Fig. 4b, Extended Data Tables 2 and 3) is reminiscent of the patterns found in experiments on the mouse (Fig. 4a). Furthermore, this relationship between talonid height and morphology was retained when we replaced cusp number with talonid complexity using orientation patch count (OPC, Fig. 4c). In contrast, the trigonid morphology remained relatively constrained (Extended Data Table 2). OPC, which is calculated as the number of discrete surfaces distinguished by differences in

b, Talonid height characters are tabulated as the height of the talonid (white arrowhead) relative to the trigonid (black arrowhead). **c**, Differential sensitivities of main regions of tooth crown to EDA show how different parts of the crown have varying trait sensitivities to EDA. For the full range of data, see Supplementary Table 1.

orientation, has been shown to increase across the dietary spectrum from carnivores to omnivores to herbivores in extant mammals^{27,28}. Therefore, even though carnivoran dental diversity is driven by ecological and functional factors, development may have an influence on which parts undergo adaptations more easily.

Taken together, the same developmental cascade, starting from the trigonid, may have contributed both to the initial evolution of the talonid at the base of mammalian evolution and to dental morphological diversification during mammalian radiations. To a lesser degree, the same pattern may hold for the anterior end of the teeth, and it is conceivable that an analogous developmental cascade to the one that produced the talonid also produced the anterior expansion of the crown in pseudo-tribosphenic mammals²⁹. A cascading system of activation–inhibition between teeth has been proposed to regulate molar proportions in mammals³⁰, and in general, much of the evolutionary history of mammalian dentitions may consist of tinkering³¹ with this pre-existing developmental program.

Retrieving ancestral character states

Despite the overall agreement between experimental and evolutionary patterns, there were small but important differences when compared with the details of rodent evolution. Most notably, the *Eda* null teeth had fused cusps, which differs from what is found in basal rodentiaforms

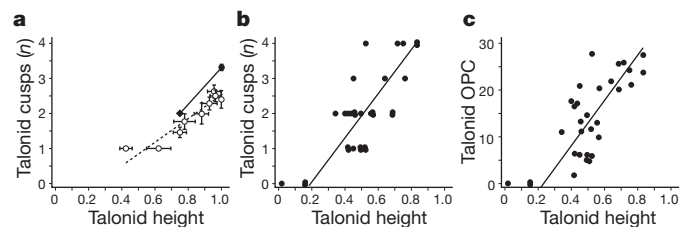


Figure 4 | Testing developmental predictions on evolutionary patterns.

a, Relative height of the m1 talonid (measured as percentage of the trigonid height) in each treatment (open circles, error bars denote s.e.m.) correlates with the number of cusps in the talonid. The experimental data bridges the corresponding values for murine rodent species ($n = 35$, black circle with s.e.m.) and for a basal rodentiaform *Tribosphenomys minutus* (black diamond). The reduced major axis regression slope for the experimental data including untreated *Eda* null and wild-type teeth is 3.30 and the intercept is -1.79 ($r^2 = 0.891$) and the corresponding values for only explants with EDA are 4.22 and -1.63 ($r^2 = 0.946$). We note that these slopes can be considered underestimates due to the variably present hypoconulid cusps in cultured wild-type mouse teeth. **b**, **c**, The first lower molars of carnivoran species ($n = 32$) show correlated changes in the talonid height and cusp number (**b**) and the talonid height and talonid complexity (**c**) measured using OPC. The reduced major-axis regression slope for the graph in **b** is 6.53 and the intercept is -1.38 ($r^2 = 0.594$), and for the graph in **c** it is 48.24 and the intercept is -10.93 ($r^2 = 0.562$). For data and details see Methods and Extended Data Tables 1–3.

and close relatives^{2–4}. This difference in morphology indicated that, in addition to EDA signalling, other pathways needed to be adjusted in order to produce the evolutionarily basal morphologies. To address this issue, we considered reducing the required cusp spacing, which can be experimentally adjusted beyond the normal mouse pattern by modulating multiple signalling pathways¹⁹. We therefore next set out to engineer mouse teeth that would have additional characters of basal rodents (see Methods).

First, we modelled the reduction in the cusp spacing by decreasing inhibition in the simulated *Eda* null teeth, which resulted in formation of multiple cusps (Extended Data Fig. 5). Next, because SHH has been shown to inhibit cusp formation by regulating cusp spacing^{19,32}, we cultured *Eda* null samples with a SHH inhibitor, thereby inhibiting the inhibitor. This treatment also circumvents the tendency of EDA to cause the formation of crests or lophs between cusps^{10,19,33}, which are found in evolutionarily derived rodents. The experimental results validated the *in silico* model simulations: inhibition of SHH signalling in cultured *Eda* null teeth caused the development of more distinct cusps, without eliminating evolutionarily basal features of *Eda* null teeth such as the height difference between trigonid and talonid (Fig. 5a, Extended Data Fig. 6).

To push the experimental system further and to retrieve features present in the basal taxon *Tribosphenomys*², we inhibited SHH in developing *Eda* null teeth *in vivo* (see Methods). *Tribosphenomys* cusps are columnar and well separated, lacking a crest called the metalophid (the trigonid wall) that connects cusps together. The loss of the metalophid has been linked to the basal rodentiaforms², but it has reappeared in many rodent clades, including murines^{2,3,24,25}. Our *in vivo* engineered tooth shapes showed columnar and laterally separated cusps without the metalophid

(Fig. 5b), a morphology visible also at the enamel–dentin junction (Extended Data Fig. 7). Although the effects of SHH inhibition were variable, the trough separating the protoconid and metaconid approximated the pattern found in *Tribosphenomys* (Fig. 5c). These results indicate that, with a relatively small number of developmental changes, mouse teeth can be engineered to express evolutionarily basal traits.

Conclusions

Our results demonstrate that many of the step-wise transitions that are widespread in the fossil record of mammalian teeth are reproducible experimentally. Whereas our results suggest that several, if not the majority, of dental traits are developmentally linked¹⁰, individual characters may respond to different levels of the same signal. These thresholds may well underlie different morphological gradients that have been identified along the tooth row³⁴. Moreover, trait thresholds may affect evolutionary rates of individual traits differently. Such data in turn should be useful when combined with analyses of character correlations³⁵, and in weighting or ordering characters, or objectively assigning transition weights within characters coded to minimize character dependency effects in phylogenetic analyses⁸. Other developmental factors and signalling pathways may influence traits in other ways, but we predict that the general pattern of results will hold as long as the factors affect the signalling dynamics of the enamel knots. Finally, as recently proposed for the evolution of bird and non-avian dinosaur skulls³⁶, developmental data can suggest novel insights into the processes underlying heterochrony. A better mechanistic basis for heterochrony will help to explain changes in evolutionary rates, including prediction of intermediate morphologies even when they have not yet been recovered in the fossil record. In general, with advancing understanding of development, it will be possible to experimentally test many more of the known evolutionary transitions.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 March; accepted 25 June 2014.

Published online 30 July 2014.

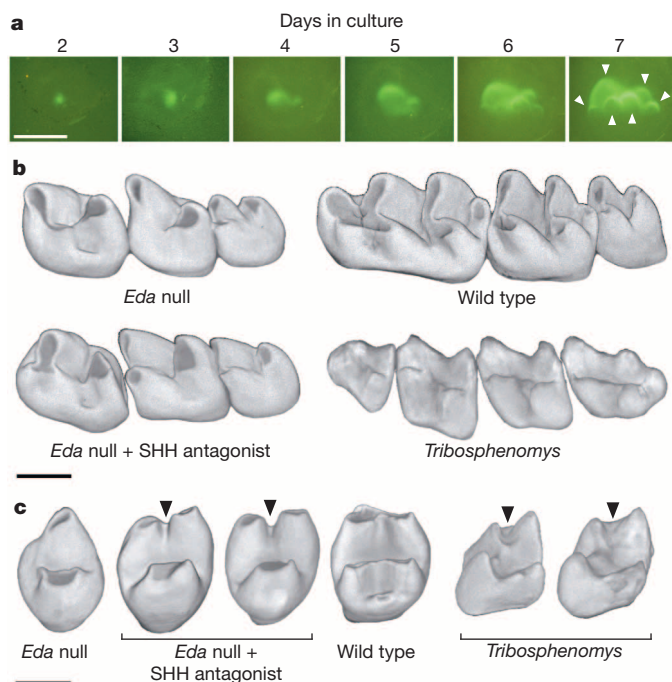


Figure 5 | Engineering mouse teeth to have basal character states. **a**, *Eda* null teeth cultured with SHH antagonist ($n = 9$ of 11 teeth) show more and better separated cusps (arrowheads, compare to Fig. 1a). **b**, Second molars produced using *in vivo* treatment of *Eda* null mice with SHH antagonist particularly show better separation of cusps compared to the *Eda* null teeth ($n = 2$ of 4 mice). *Tribosphenomys minutus* teeth lack crests connecting cusps (specimen V10776 shows p4–m3). **c**, Obliquely posterior views of molars show the lack of the metalophid crest (arrowheads) in treated *Eda* null, which has replicated the ancestral morphology of *T. minutus*. The *Tribosphenomys* molars shown are the first (V10776, on the left) and the second molar (V10775 holotype, on the right). Teeth have been mirrored if needed to represent the left side, anterior is towards the left in **a** and **b** and top in **c**. Scale bars, 500 μm .

- Luo, Z. X., Cifelli, R. L. & Kielan-Jaworowska, Z. Dual origin of tribosphenic mammals. *Nature* **409**, 53–57 (2001).
- Meng, J. & Wyss, A. R. The morphology of *Tribosphenomys* (Rodentiaformes, Mammalia): Phylogenetic implications for basal Glires. *J. Mamm. Evol.* **8**, 1–71 (2001).
- Asher, R. J. *et al.* Stem lagomorpha and the antiquity of Glires. *Science* **307**, 1091–1094 (2005).
- O’Leary, M. A. *et al.* The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662–667 (2013).
- Kluge, A. G. & Farris, J. S. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* **18**, 1–32 (1969).
- Felsenstein, J. Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst. Zool.* **22**, 240–249 (1973).
- Doyle, J. J. Trees within trees: genes and species, molecules and morphology. *Syst. Biol.* **46**, 537–553 (1997).
- O’Keefe, F. R. & Wagner, P. J. Inferring and testing hypotheses of cladistic character dependence by using character compatibility. *Syst. Biol.* **50**, 657–675 (2001).
- Wake, D. B. Phylogenetic implications of ontogenetic data. *Geobios* **22** (suppl. 2), 369–378 (1989).
- Kangas, A. T., Evans, A. R., Thesleff, I. & Jernvall, J. Nonindependence of mammalian dental characters. *Nature* **432**, 211–214 (2004).
- Luo, Z.-X. Transformation and diversification in early mammal evolution. *Nature* **450**, 1011–1019 (2007).
- Polly, P. D. Developmental dynamics and G-matrices: can morphometric spaces be used to model evolution and development? *Evol. Biol.* **35**, 83–96 (2008).
- Rice, S. H. Theoretical approaches to the evolution of development and genetic architecture. *Ann. NY Acad. Sci.* **1133**, 67–86 (2008).
- Salazar-Ciudad, I. & Marin-Riera, M. Adaptive dynamics under development-based genotype-phenotype maps. *Nature* **497**, 361–364 (2013).
- Grüneberg, H. Genes and genotypes affecting the teeth of the mouse. *J. Embryol. Exp. Morphol.* **14**, 137–159 (1965).
- Harfe, B. D. *et al.* Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
- Jernvall, J., Keränen, S. V. E. & Thesleff, I. Evolutionary modification of development in mammalian teeth: quantifying gene expression patterns and topography. *Proc. Natl Acad. Sci. USA* **97**, 14444–14448 (2000).

18. Gaide, O. & Schneider, P. Permanent correction of an inherited ectodermal dysplasia with recombinant EDA. *Nature Med.* **9**, 614–618 (2003).
19. Harjunmaa, E. *et al.* On the difficulty of increasing dental complexity. *Nature* **483**, 324–327 (2012).
20. Arthur, W. *Evolution: a Developmental Approach* (Wiley-Blackwell, 2011).
21. Salazar-Ciudad, I. & Jernvall, J. A computational model of teeth and the developmental origins of morphological variation. *Nature* **464**, 583–586 (2010).
22. Butler, P. M. Early trends in the evolution of tribosphenic molars. *Biol. Rev. Camb. Philos. Soc.* **65**, 529–552 (1990).
23. Osborn, J. W. & Crompton, A. W. The evolution of mammalian from reptilian dentitions. *Breviora* **399**, 1–18 (1973).
24. Wessels, W. Miocene rodent evolution and migration: Muroidea from Pakistan, Turkey and North Africa. *Geol. Ultraiectina* **307**, 1–290 (2009).
25. López Antoñanzas, R. First *Potwarmus* from the Miocene of Saudi Arabia and the early phylogeny of murines (Rodentia: Muroidea). *Zool. J. Linn. Soc.* **156**, 664–679 (2009).
26. Van Valkenburgh, B. Major patterns in the history of carnivorous mammals. *Annu. Rev. Earth Planet. Sci.* **27**, 463–493 (1999).
27. Evans, A. R., Wilson, G. P., Fortelius, M. & Jernvall, J. High-level similarity of dentitions in carnivores and rodents. *Nature* **445**, 78–81 (2007).
28. Santana, S. E., Strait, S. & Dumont, E. R. The better to eat you with: functional correlates of tooth structure in bats. *Funct. Ecol.* **25**, 839–847 (2011).
29. Luo, Z. X., Ji, Q. & Yuan, C. X. Convergent dental adaptations in pseudo-tribosphenic and tribosphenic mammals. *Nature* **450**, 93–97 (2007).
30. Kavanagh, K. D., Evans, A. R. & Jernvall, J. Predicting evolutionary patterns of mammalian teeth from development. *Nature* **449**, 427–432 (2007).
31. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
32. Cho, S.-W. *et al.* Interactions between Shh, Sostdc1 and Wnt signaling and a new feedback loop for spatial patterning of the teeth. *Development* **138**, 1807–1816 (2011).
33. Gomes Rodrigues, H. G. *et al.* Roles of dental development and adaptation in rodent evolution. *Nat. Commun.* **4**, 2504 (2013).
34. Van Valen, L. An analysis of developmental fields. *Dev. Biol.* **23**, 456–477 (1970).
35. Goswami, A. & Polly, P. D. in *Carnivoran Evolution: New Views on Phylogeny, Form, and Function* (eds Goswami, A. & Friscia, A.) 141–164 (Cambridge Univ. Press, 2010).
36. Bhullar, B.-A. *et al.* Birds have paedomorphic dinosaur skulls. *Nature* **487**, 223–226 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Fortelius, J. Eronen, I. Thesleff, P. Munne for discussions; S. Alto, M. Mäkinen, R. Santalahti, R. Savolainen, and M. Christensen for technical assistance; P. Schneider for the Fc-EDA-A1-protein; F. de Sauvage for HhAntag compound; Hou Yemao for tomography; and the Finnish Museum of Natural History (Helsinki, Finland), Museum of Natural History (Stockholm, Sweden), and Museum für Naturkunde (Berlin, Germany) for specimen loans. This work was supported by the Academy of Finland to J.J., M.M., I.S.-C., R01-DE021420 (NIH/NIDCR) and an NIH Director's New Innovator Award DP2-OD007191 to O.D.K., an Australian Research Council Future Fellowship to A.R.E. and the Major Basic Research Projects (2012CB821904) of MST to Z.-Q.Z. Data are presented in the Supplementary and Extended Data Tables, available in the MorphoBrowser database (<http://morphobrowser.biocenter.helsinki.fi/>) and from the authors, and models can be accessed at (<http://www.biocenter.helsinki.fi/bi/evodevo/toothmaker.html>).

Author Contributions E.H., J.J. and O.D.K. designed the project and wrote the initial manuscript. E.H. and E.R. performed culturing experiments and K.S. mouse experiments. E.H., E.R., A.K. and J.J. performed measurements and prepared images. I.J.C., A.R.E. and J.J. analysed evolutionary data. I.S.-C. constructed the computational model and T.H. the ToothMaker. M.L.M., Z.-Q.Z. provided materials, observations and scientific interpretations. O.D.K. and J.J. coordinated the study. All authors discussed the results and provided input on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.J. (jernvall@fastmail.fm) or O.D.K. (ophir.klein@ucsf.edu).

Structure of the DDB1–CRBN E3 ubiquitin ligase in complex with thalidomide

Eric S. Fischer^{1,2}, Kerstin Böhm^{1,2}, John R. Lydeard³, Haidi Yang⁴, Michael B. Stadler^{1,2,5}, Simone Cavadini^{1,2}, Jane Nagel⁴, Fabrizio Serluca⁴, Vincent Acker⁶, Gondichatnahalli M. Lingaraju^{1,2}, Ritesh B. Tichkule⁴, Michael Schebesta⁴, William C. Forrester⁴, Markus Schirle⁴, Ulrich Hassiepen⁶, Johannes Ottl⁶, Marc Hild⁴, Rohan E. J. Beckwith⁴, J. Wade Harper³, Jeremy L. Jenkins⁴ & Nicolas H. Thomä^{1,2}

In the 1950s, the drug thalidomide, administered as a sedative to pregnant women, led to the birth of thousands of children with multiple defects. Despite the teratogenicity of thalidomide and its derivatives lenalidomide and pomalidomide, these immunomodulatory drugs (IMiDs) recently emerged as effective treatments for multiple myeloma and 5q-deletion-associated dysplasia. IMiDs target the E3 ubiquitin ligase CUL4–RBX1–DDB1–CRBN (known as CRL4^{CRBN}) and promote the ubiquitination of the IKAROS family transcription factors IKZF1 and IKZF3 by CRL4^{CRBN}. Here we present crystal structures of the DDB1–CRBN complex bound to thalidomide, lenalidomide and pomalidomide. The structure establishes that CRBN is a substrate receptor within CRL4^{CRBN} and enantioselectively binds IMiDs. Using an unbiased screen, we identified the homeobox transcription factor MEIS2 as an endogenous substrate of CRL4^{CRBN}. Our studies suggest that IMiDs block endogenous substrates (MEIS2) from binding to CRL4^{CRBN} while the ligase complex is recruiting IKZF1 or IKZF3 for degradation. This dual activity implies that small molecules can modulate an E3 ubiquitin ligase and thereby upregulate or downregulate the ubiquitination of proteins.

Thalidomide (α -(*N*-phthalimido)glutaramide) was introduced to the market in 1954 by the company Chemie Grünenthal. Initially promoted as a sedative with anti-emetic properties^{1,2}, it became popular for treating ‘morning sickness’³. In 1961, thalidomide taken in the first trimester of pregnancy was implicated in frequent limb deformities in infants^{4,5}. Between 8,000 and 12,000 affected children were born before the drug was banned. Interest in thalidomide revived in 1965, when it was shown to have immunomodulatory and anti-inflammatory properties in patients with erythema nodosum leprosum, an inflammatory complication of leprosy⁶. In 1994, thalidomide was found to inhibit the basic fibroblast growth factor (bFGF)-induced formation of new blood vessels⁷. These findings prompted clinical trials exploring thalidomide use for anti-angiogenic cancer therapy. The efficacy of thalidomide and its derivatives lenalidomide and pomalidomide (collectively known as IMiDs) has since been demonstrated for several haematological cancers⁸: newly diagnosed multiple myeloma (thalidomide)⁹, refractory multiple myeloma (lenalidomide or pomalidomide) and 5q-deletion-associated myelodysplastic syndrome (lenalidomide).

The target of thalidomide, cereblon (CRBN), is a ubiquitously expressed protein that is part of the cullin-4-containing E3 ubiquitin ligase complex CUL4–RBX1–DDB1 (known as CRL4)¹⁰. Mutations in CRBN are associated with autosomal recessive, non-syndromic mental retardation¹¹. In myeloma cells, the anti-proliferative activities of IMiDs are linked to CRBN expression^{12,13}, making IMiDs the first clinically approved drug targeted at E3 ubiquitin ligases with specificity for CUL4–RBX1–DDB1–CRBN (CRL4^{CRBN})¹². The anti-proliferative and immunomodulatory effects of IMiDs have recently been linked to drug-induced ubiquitination and degradation of the transcription factors IKZF1 (also known as IKAROS) and IKZF3 (also known as AIOLOS) by CRL4^{CRBN} (refs 14–16). Accordingly, loss of CRBN is a common determinant of drug resistance in myeloma cells¹². How IMiD binding affects CRL4^{CRBN} at the molecular level remains unclear. We set out to examine the role of CRBN within the E3 ubiquitin

ligase complex CRL4^{CRBN}, characterizing the effect of IMiD binding on ligase activity.

Structure of DDB1–CRBN bound to IMiDs

We crystallized a chimaeric complex of human DDB1 and *Gallus gallus* (chicken) CRBN bound to thalidomide (refined to 3.0 Å), lenalidomide (3.0 Å) and pomalidomide (3.5 Å) (Fig. 1a, b and Extended Data Table 1). The high level of sequence conservation between human and chicken CRBN (Extended Data Fig. 1a, b) allows structural insight into human CRBN to be inferred directly from *G. gallus* CRBN. All subsequent biochemical and cell-biological experiments were performed with full-length human proteins. *G. gallus* CRBN consists of three sub-domains (Extended Data Fig. 2a–f): a seven-stranded β -sheet located in the amino-terminal domain (NTD, residues 1–185) (Extended Data Fig. 2a), a seven α -helical bundle domain (HBD, residues 186–317) involved in DDB1 binding (Fig. 1c), and a carboxy-terminal domain composed of eight β -sheets (CTD, residues 318–445) (Fig. 1b). DDB1 comprises three seven-bladed WD40 β -propellers (BPA, BPB and BPC) arranged in a triangular fashion¹⁷, with *G. gallus* CRBN attaching to a cavity between the BPA and BPC propellers (Fig. 1c). The molecular basis of the HBD-mediated attachment of *G. gallus* CRBN to DDB1 defines a novel class of DDB1 binders and differs in detail from previously reported DDB1 attachment modules^{17–20} (Extended Data Fig. 2e, f).

The *G. gallus* CRBN N-terminal region (residues 46–317), including the NTD and HBD, resembles the N-terminal domain of bacterial Lon proteases (Protein Data Bank (PDB) ID 3LJC; root mean squared deviation (r.m.s.d.), 2.7 Å over 178 residues aligned) (Extended Data Fig. 2b). The CTD harbours the thalidomide-binding pocket and contains a conserved Zn²⁺-binding site, which is situated approximately 18 Å from the IMiD (Fig. 1a, b). The Zn²⁺ ion is coordinated through conserved cysteine residues 325, 328, 393 and 396. The *G. gallus* CRBN–CTD shares structural

¹Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. ²University of Basel, Petersplatz 10, CH-4003 Basel, Switzerland. ³Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA. ⁴Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. ⁵Swiss Institute of Bioinformatics, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. ⁶Novartis Pharma AG, Institutes for Biomedical Research, Novartis Campus, CH-4056 Basel, Switzerland.

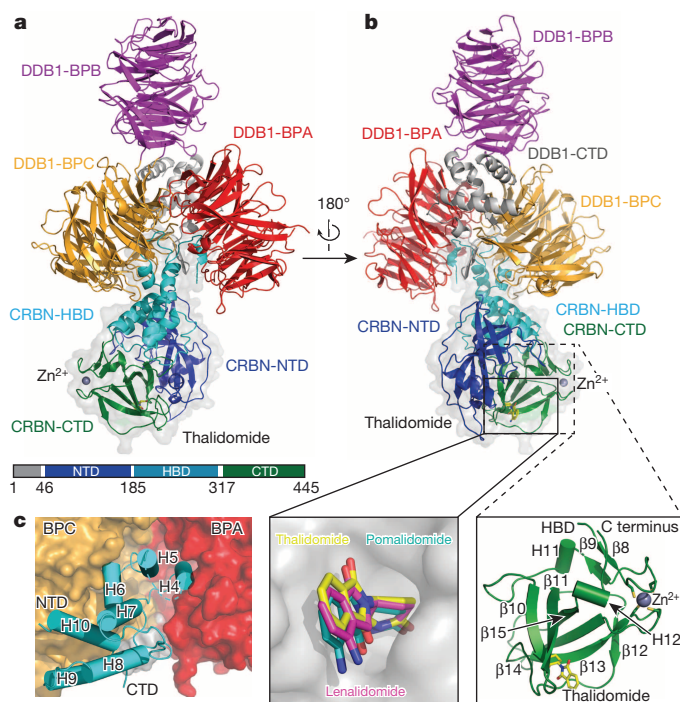


Figure 1 | The overall structure of the DDB1-CRBN complex. **a**, Cartoon representation of the structure of the complex of human DDB1, *G. gallus* CRBN and thalidomide: DDB1, highlighting the domains BPA (red), BPB (magenta), BPC (orange) and DDB1-CTD (grey); *G. gallus* CRBN, highlighting the domains NTD (blue), HBD (cyan) and CTD (green). The Zn^{2+} ion is drawn as a grey sphere. **b**, As in **a**, with thalidomide shown as a yellow stick structure. A close-up showing that all IMiDs occupy a common binding site on CRBN (solid boxed area; red, oxygen; blue, nitrogen) and a close-up of the overall *G. gallus* CRBN-CTD architecture (dashed boxed area) are shown. **c**, *G. gallus* CRBN-HBD helices and their interactions with DDB1.

similarity with the pseudouridine synthase and archaeosine transglycosylase (PUA) fold family²¹, which are involved in binding diverse sets of ligands (Extended Data Fig. 2c, d).

IMiD binding to CRBN

Thalidomide, lenalidomide and pomalidomide (Fig. 2a–c and Extended Data Fig. 3a–i) bind a pocket on *G. gallus* CRBN-CTD (Fig. 1b) situated in a surface groove that is highly conserved across CRBN orthologues (Extended Data Fig. 1b). The three ligands superimpose with very little deviation in the α -(isoidolinone-2-yl) glutarimide moiety, which contributes the majority of interactions between the receptor and the compounds and is the main pharmacophore²². The glutarimide group is held in a buried cavity between *G. gallus* CRBN sheets β 10 and β 13 (Fig. 2d). The glutarimide carbonyls (C2 and C6) and the intervening amide (N1) are in hydrogen-bonding distance with *G. gallus* CRBN residues His 380 and Trp 382, respectively (Fig. 2c, d). A delocalized lone pair connects the glutarimide nitrogen with the two glutarimide carbonyls (C2–N1–C6) and is coplanar with Trp 382. The opposing aliphatic face of the glutarimide ring (C3, C4 and C5) is in tight van der Waals contact with a hydrophobic pocket lined by Trp 382, Trp 388, Trp 402 and Phe 404. *In vitro*, mutations of Tyr 386 (which affect the integrity of the binding pocket) and Trp 388 (which is directly involved in compound binding) to alanine ablate the binding of all three IMiDs to CRBN¹⁰ (Extended Data Fig. 4a–c). Mutations of the equivalent residues render CRL4^{CRBN} insensitive to the presence of thalidomide or lenalidomide *in vivo*^{10,12}. In addition to the almost identical binding modes, we found that thalidomide, lenalidomide and pomalidomide have similar affinities for CRBN, with dissociation constants of ~ 250 nM, ~ 178 nM and ~ 157 nM, respectively (Extended Data Fig. 4d–h).

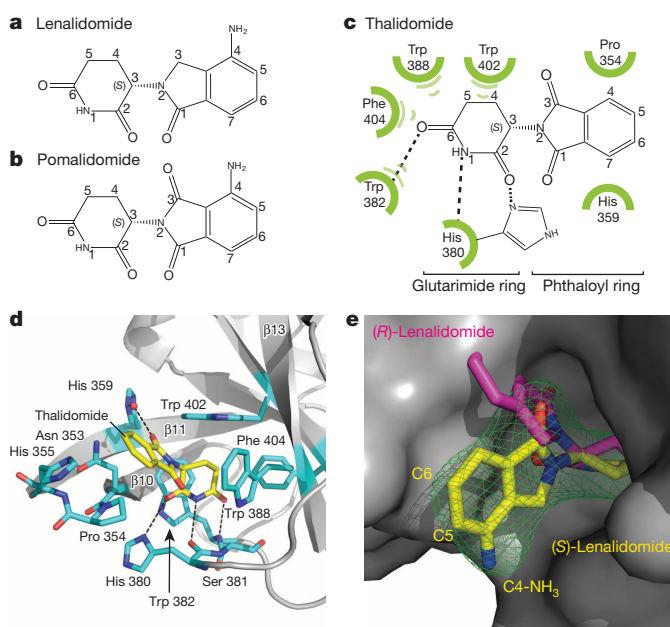


Figure 2 | IMiD binding to CRBN. **a**, Chemical structure of lenalidomide. **b**, Chemical structure of pomalidomide. **c**, Sketch of thalidomide and its interactions with *G. gallus* CRBN. Hydrogen bonds are shown as dashed lines, and hydrophobic interactions are indicated as green semicircles. **d**, IMiDs are anchored through hydrogen bonding of the glutarimide moiety to *G. gallus* CRBN His 380 and Trp 382, as well as through the aliphatic face of the glutarimide being engulfed in a hydrophobic cage. **e**, Surface representation of *G. gallus* CRBN (grey) and (S)-lenalidomide shown as a yellow stick structure, together with its positive $mF_o - DF_c$ electron density map ($\sigma = 3.5$) shown in green. The fit of the (S)- and (R)-enantiomers is indicated.

Thalidomide differs from lenalidomide and pomalidomide in the C4 phthalimide aniline functionality (Fig. 2a–c), which is found in a solvent-exposed position. The common carbonyl at the phthaloyl C1 position contributes a water-mediated hydrogen bond to His 359, which anchors the phthaloyl ring system together with stacking interactions provided by the aliphatic face of Pro 354 (Fig. 2d). The phthalimide C5 and C6 positions are fully solvent exposed. The overall shape of the buried IMiD-binding pocket favours binding of the (S)-enantiomer over the (R)-enantiomer (Fig. 2e), which is in agreement with *in vivo* experiments¹².

CRBN functions as a DCAF for the ligase CRL4^{CRBN}

Within the CRL4 family of ligases, DDB1 functions as the adaptor connecting the substrate receptor to the ligase CUL4 (refs 17, 19, 23). More than a dozen substrate receptors, including CRBN, have been identified (and these are designated DCAFs for DDB1- and CUL4-associated factors). *G. gallus* CRBN, despite lacking the canonical DCAF WD40 fold, resembles a substrate receptor in its dimensions and position on DDB1 (Extended Data Fig. 5a, b). The thalidomide-binding site is situated where substrates generally bind to WD40 DCAFs (see, for example, DDB2 engaging DNA (Fig. 3a and Extended Data Fig. 5a, b)). The equivalent residues in the structurally related PUA-domain-containing proteins are directly engaged in ligand binding (Extended Data Fig. 5c–e). The PUA domain of CRBN has striking structural similarity to a member of the methionine sulfoxide reductase family (Extended Data Fig. 2c, d). In *G. gallus* CRBN, the now defunct active centre of the reductase interacts with IMiDs. Truncation of the CRBN C terminus bordering the conserved thalidomide-binding pocket has been found in non-syndromic mental retardation (see Extended Data Fig. 6a–c for analysis of CRBN mutations).

Within CUL4–RBX1–DDB1–DCAF complexes, CUL4 was found to freely rotate up to 150° around DDB1 (refs 24, 25) (Fig. 3a, b). CUL4 mobility is undirected and is driven exclusively by Brownian motion. Given the strictly modular architecture of the CRL4 family^{23–25}, the structure

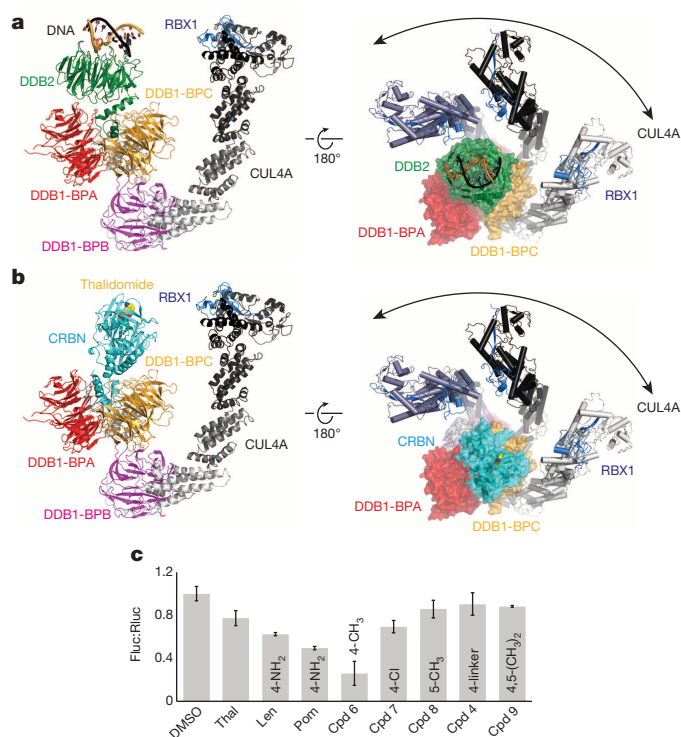


Figure 3 | CRBN is a substrate receptor in the ligase CRL4^{CRBN}. **a**, Architecture of the CRL4^{DDB2} complex bound to DNA (PDB ID 4A0K). **b**, Model of CRL4^{CRBN} bound to thalidomide. **c**, Firefly luciferase (Fluc) to Renilla luciferase (Rluc) ratios (Fluc:Rluc) of IKZF1-reporter-plasmid-transfected HEK 293T cells following incubation with the indicated thalidomide derivatives. The data are presented as mean \pm s.e.m. ($n = 4$). Cpd, compound; Len, lenalidomide; Pom, pomalidomide; Thal, thalidomide.

of CRL4^{CRBN} can be predicted with high confidence (Fig. 3b). Rotation of the ligase arm of CUL4 around DDB1 and CRBN establishes a ubiquitination zone with dimensions of up to $340 \text{ \AA} \times 110 \text{ \AA} \times 30 \text{ \AA}$ (Fig. 3b), with the centre of rotation near the thalidomide-binding site. The CRL4 ligase is promiscuous, targeting lysines that cross this ubiquitination zone. Accordingly, we observed that CRBN is autoubiquitinated *in vitro* (Extended Data Fig. 7a–d) at the unstructured N-terminal tail²⁶ (residues Lys 39 and Lys 43; see also Extended Data Fig. 7d). We found that CRBN autoubiquitination persisted in the presence of IMiDs *in vitro* (Extended Data Fig. 7b) and was subject to inhibition by the Cop9 signalosome (CSN) (Extended Data Fig. 7b, c).

Pomalidomide and lenalidomide effectively target the IKAROS transcription factors IKZF1 and IKZF3 for degradation by CRL4^{CRBN}. Thalidomide, by contrast, is here less efficient^{14–16} (Fig. 3c). All three IMiDs have comparable affinities for CRBN (Extended Data Fig. 4e–h) and similar predicted membrane permeabilities (Extended Data Fig. 7e). The major structural difference between CRBN-bound lenalidomide and pomalidomide and CRBN-bound thalidomide lies in the presence of the solvent-exposed C4 aniline functionality in lenalidomide and pomalidomide. Therefore, different functional groups at the phthaloyl C4, C5 and C6 positions were tested for their ability to degrade IKZF1 in a dual luciferase reporter assay¹⁴ (see Supplementary Methods). Small groups attached to C4 of thalidomide (NH_2 , CH_3 and to some extent Cl) promoted IKZF1 degradation (Fig. 3c). By contrast, larger substituents at the C4, C5, or C4 and C6 positions were less effective at promoting IKZF1 degradation. These modifications are not expected to affect CRBN binding, as even a large substituent at the C4 position had no adverse effect on affinity (Extended Data Fig. 4d). Assuming comparable cellular concentrations^{27–29}, this finding indicates that solvent-exposed bulky groups at C4 (and methyl groups at C5 and C6) interfere with IKZF1 binding. The aniline and methyl substituents at C4, however, appear to contribute to IKZF1 degradation, probably through direct interactions.

MEIS2 is an endogenous ligase substrate of CRL4^{CRBN}

Next, we set out to examine the effects of IMiDs on endogenous CRBN substrates, which have so far remained elusive. An unbiased biochemical screen was established to identify proteins that are ubiquitinated by CRL4^{CRBN}. Human protein microarrays (~9,000 proteins) were used for on-chip ubiquitination by the complex CUL4A–RBX1–DDB1–CRBN (CRL4^{CRBN}) in the presence of E1 (UBA1), E2 (UBCH5A), ubiquitin (biotin–ubiquitin) and ATP (Extended Data Fig. 8a–e). We reasoned that a substrate would be subject to ubiquitination by CRL4^{CRBN} but not by a control ligase (CRL4^{CDT2}), that a substrate would overcome inhibition of CRL4^{CRBN} by CSN²⁴ and that ubiquitination of such substrates would be inhibited by lenalidomide. Following cluster analysis (43 clusters), we identified one cluster that best matched the expected ubiquitination profile (see Supplementary Methods for details). In an orthogonal screen, the top five candidate genes (*GRINL1A*, *MBOAT7*, *OTUD7B*, *C6orf141* and *MEIS2*) were overexpressed in HEK 293T cells and assessed for lenalidomide-induced changes in steady-state levels (Extended Data Fig. 8f). Of these factors, we focused on MEIS2, a transcription factor that has been implicated in various aspects of human development^{30–32}, which we found was stabilized on lenalidomide treatment (see below).

To recapitulate the on-chip results in solution, insect-cell-derived MEIS2 was used to establish a fully recombinant CRL4^{CRBN} ubiquitination system. MEIS2 expressed from insect cells was isolated in a hyper-phosphorylated form. Dephosphorylation was found to improve protein behaviour, resulting in increased MEIS2 ubiquitination using lysine-free (K0) ubiquitin (Fig. 4a, compare lanes 1 and 2). MEIS2 ubiquitination was also observed using phosphorylated MEIS2 (data not shown). MEIS2 ubiquitination by CRL4^{CRBN} was subject to inhibition by various IMiDs (Fig. 4a, lanes 4–8, and Extended Data Fig. 9a), irrespective of the chemical substituent at the C4, C5 or C6 position. A CRL4^{CRBN} mutant carrying the Tyr384Ala and Trp386Ala substitutions (CRL4^{YW/AA}) that rendered the receptor unable to bind IMiDs (Extended Data Fig. 4a–c) also impaired

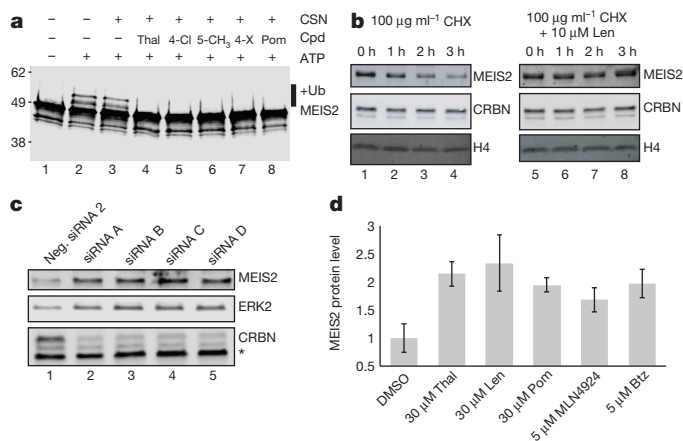


Figure 4 | MEIS2 is an endogenous substrate of the ligase CRL4^{CRBN}. **a**, CRL4^{CRBN} ubiquitinates MEIS2 *in vitro* (lane 2), a reaction inhibited by thalidomide (Thal, lane 4), pomalidomide (Pom, lane 8), compound (Cpd) 7 (lane 5), compound 8 (lane 6) and compound 4 (X, extended carbon linker) (lane 7). Lysine-free (K₀) ubiquitin (Ub) was used to obtain a defined band-shift detected by anti-MEIS2 immunoblotting. **b**, SK-N-DZ cells were pretreated with 10 μM lenalidomide or DMSO before addition of 100 μg ml⁻¹ CHX for the indicated times. MEIS2 and CRBN levels were detected using anti-MEIS2 and anti-CRBN immunoblotting. Histone H4 served as a loading control. **c**, M059J cells were transfected with one of four short interfering RNA (siRNA) constructs (labelled A to D) targeting CRBN, or a negative control (Neg. siRNA2), and the levels of endogenous CRBN and MEIS2 proteins were monitored. The asterisk indicates a non-specific band, and ERK2 served as a loading control. **d**, Treatment of M059J cells with the indicated amounts of IMiDs, MLN4924 or bortezomib (Btz) resulted in an increase in the steady-state MEIS2 protein levels after 4 h. The data are presented as mean \pm s.e.m. (IMiDs, $n = 3$; MLN4924 and Btz, $n = 2$).

MEIS2 ubiquitination (Extended Data Fig. 9b). In accordance with our on-chip results, we found that MEIS2 can relieve the inhibition of CRL4^{CRBN} by CSN (Extended Data Fig. 9c) and is not ubiquitinated by the control ligases CRL4^{CSA} and CRL4^{CDT2} (Extended Data Fig. 9d). These data suggest that, within the CRL4^{CRBN} complex, CRBN targets MEIS2 for ubiquitination *in vitro* with the aid of its IMiD-binding site.

We next sought to test the effect of lenalidomide on MEIS2 half-life in cells by performing cycloheximide (CHX) chase experiments. RNA-seq data from The Cancer Genome Atlas (TCGA) identified the neuroblastoma cell line SK-N-DZ as having high levels of endogenous MEIS2 and CRBN transcripts. Following treatment with 100 µg ml⁻¹ CHX, we found that MEIS2 protein levels were largely depleted after 3 h (Fig. 4b, lanes 1–4). The addition of 10 µM lenalidomide stabilized MEIS2 protein levels under these conditions (Fig. 4b, lanes 5–8). As SK-N-DZ cells were recalcitrant to transfection, we subsequently employed M059J cells, which are similarly characterized by high levels of MEIS2 messenger RNA (TCGA), to examine the effects of CRBN depletion by RNA interference on MEIS2 abundance. For all of the experiments, the endogenous MEIS2 levels were monitored by quantitative immunoblotting using infrared detection (see Extended Data Fig. 9e–p). M059J or HEK 293T cells were transfected with four short interfering RNAs, resulting in efficient CRBN knockdown (Fig. 4c) and increased MEIS2 steady-state levels in M059J cells (Fig. 4c) and HEK 293T cells (data not shown), implicating CRBN in MEIS2 turnover.

To test whether lenalidomide treatment results in increased steady-state levels of MEIS2, M059J cells were treated with lenalidomide or a dimethyl sulphoxide (DMSO) control. Following lenalidomide treatment, endogenous MEIS2 protein levels were elevated by up to 2.5-fold (Fig. 4d and Extended Data Fig. 9i–k). A similar level of steady-state MEIS2 stabilization was observed with MLN4924 (Extended Data Fig. 9i, lane 6), a NEDD8-activating enzyme inhibitor, and with the proteasome inhibitor bortezomib (Extended Data Fig. 9i, lane 5). Increases in MEIS2 protein levels following lenalidomide exposure were not due to mRNA changes, as determined by quantitative reverse transcription PCR (RT-PCR) (Extended Data Fig. 9m).

When examining different IMiD derivatives, we found that thalidomide, lenalidomide and pomalidomide stabilized steady-state MEIS2 protein levels to a similar extent (Fig. 4d and Extended Data Fig. 9i, k). The effect of thalidomide on MEIS2 levels was also observed *in vivo*: whole zebrafish embryos 24 h post fertilization showed a 1.5-fold increase in MEIS2 levels, comparable to that observed in the cell lines (Extended Data Fig. 9l).

Thalidomide is an agonist and an antagonist

We found that IMiD binding and MEIS2 recruitment were mutually exclusive *in vitro* and *in vivo* (Figs 4a and 5a–c and Extended Data Fig. 9b). Accordingly, a stable cell line overexpressing an epitope-tagged MEIS2 together with the CRBN^{YW/AA} mutant (Extended Data Fig. 9f, g) exhibited no changes in MEIS2 levels following the addition of 40 µM lenalidomide. MEIS2 is thus an endogenous, IMiD-sensitive target of CRBN (see Supplementary Discussion for the putative role of MEIS2 in IMiD-mediated teratogenicity). Most CRLs studied so far target multiple substrates^{33,34}. Given that IMiDs occupy the canonical ligand interface of the CRBN PUA domain, we also expect that other endogenous substrates are prevented from CRBN binding by thalidomide and its derivatives. By contrast, IMiD-dependent targeting of the IKAROS transcription factors (Fig. 5d) is facilitated by specific solvent-exposed functionalities that are not involved in CRBN binding. We propose a model in which the interaction surface used for IKAROS family member binding comprises CRBN (in its IMiD-bound form) and the C4, C5 and C6 phthaloyl positions of the IMiD. While structural studies on CRBN–IMiD–IKAROS complexes are necessary to shed light on the details of the binding mode, conceptually this mechanism bears striking similarities to auxin-induced degradation of members of the Aux/IAA repressor family by the ligase TIR1 (ref. 35) and to cyclosporin-A-induced cyclophilin binding (or FK506-induced FKBP12 binding) to calcineurin^{36,37}. The CRL4 architecture

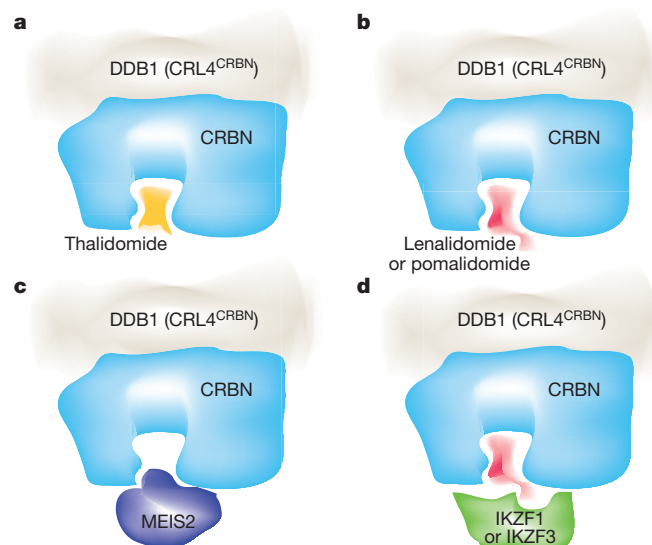


Figure 5 | Molecular model of IMiD function. **a**, Thalidomide binds to CRBN at the canonical substrate-binding site. **b**, The potent anti-myeloma drug thalidomide and its derivatives lenalidomide and pomalidomide occupy the same site but with different solvent-exposed moieties. **c**, Binding of the endogenous substrate MEIS2 and the IMiDs to this site is mutually exclusive. **d**, We propose a direct interaction of the C4 amine of lenalidomide or pomalidomide with IKAROS transcription factors (IKZF1 and IKZF3). The binding of these factors probably also involves the surrounding residues of CRBN.

supports ubiquitination in its vicinity, a property that is exploited by viral proteins in recruiting cellular targets for degradation by CRL4s¹⁷. As small molecules can apparently mimic this behaviour, it will be important to explore whether synthetic small molecules can promote the degradation of other substrates that are not typically targeted by a specific ubiquitin ligase.

Our structure–function analysis indicates that IMiD-mediated IKAROS transcription factor degradation simultaneously interferes with the recruitment of endogenous substrates (such as MEIS2) to CRL4^{CRBN}. Depending on the cell type and the proteins expressed, the administration of thalidomide and its derivatives will thus simultaneously affect the levels of two groups of proteins: upregulating the endogenous substrates while decreasing the amounts of neo-substrates. Thalidomide and its IMiD derivatives give rise to pleiotropic clinical effects, ranging from anti-cancer and immunomodulatory properties to pronounced teratogenicity. Both loss of ubiquitination, as seen for MEIS2, and gain of function, as seen for IKAROS family members, and even complex synthetic combinations of these opposing changes, need to be considered as underlying causes of these diverse clinical effects.

METHODS SUMMARY

All recombinant protein complexes were produced in insect cells²⁵. The crystal structures of human DDB1 and *G. gallus* CRBN were determined by molecular replacement with a DDB1 search model and subsequent iterative model building for CRBN. *In vitro* ubiquitination assays were performed as previously described²⁵. Protein-array-based ubiquitin ligase profiling was performed using ProtoArray v5.0 (Life Technologies) and analysed with the R software package. Cell-biological experiments were performed following standard procedures. Binding assays were based on time-resolved fluorescence resonance energy transfer (TR-FRET) or fluorescence polarization methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2013; accepted 23 May 2014.

Published online 16 July 2014.

1. Bartlett, J. B., Dredge, K. & Dalgleish, A. G. The evolution of thalidomide and its IMiD derivatives as anticancer agents. *Nature Rev. Cancer* **4**, 314–322 (2004).

2. Shortt, J., Hsu, A. K. & Johnstone, R. W. Thalidomide-analogue biology: immunological, molecular and epigenetic targets in cancer therapy. *Oncogene* **32**, 4191–4202 (2013).
3. Melchert, M. & List, A. The thalidomide saga. *Int. J. Biochem. Cell Biol.* **39**, 1489–1499 (2007).
4. McBride, W. G. Thalidomide and congenital abnormalities. *Lancet* **278**, 1358 (1961).
5. Lenz, W., Pfeiffer, R. A., Kosenow, W. & Hayman, D. J. Thalidomide and congenital abnormalities. *Lancet* **279**, 45–46 (1962).
6. Sheskin, J. Thalidomide in the treatment of lepra reactions. *Clin. Pharmacol. Ther.* **6**, 303–306 (1965).
7. D'Amato, R. J., Loughnan, M. S., Flynn, E. & Folkman, J. Thalidomide is an inhibitor of angiogenesis. *Proc. Natl Acad. Sci. USA* **91**, 4082–4085 (1994).
8. Pan, B. & Lentzsch, S. The application and biology of immunomodulatory drugs (IMiDs) in cancer. *Pharmacol. Ther.* **136**, 56–68 (2012).
9. Singhal, S. *et al.* Antitumor activity of thalidomide in refractory multiple myeloma. *N. Engl. J. Med.* **341**, 1565–1571 (1999).
10. Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
11. Higgins, J. J., Pucilowska, J., Lombardi, R. Q. & Rooney, J. P. A mutation in a novel ATP-dependent Lon protease gene in a kindred with mild mental retardation. *Neurology* **63**, 1927–1931 (2004).
12. Lopez-Girona, A. *et al.* Cereblon is a direct protein target for immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia* **26**, 2326–2335 (2012).
13. Zhu, Y. X. *et al.* Cereblon expression is required for the antimyeloma activity of lenalidomide and pomalidomide. *Blood* **118**, 4771–4779 (2011).
14. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
15. Kronke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
16. Gandhi, A. K. *et al.* Immunomodulatory agents lenalidomide and pomalidomide co-stimulate T cells by inducing degradation of T cell repressors Ikaros and Aiolos via modulation of the E3 ubiquitin ligase complex CRL4(CRBN). *Br. J. Haematol.* **164**, 811–821 (2014).
17. Li, T., Chen, X., Garbutt, K. C., Zhou, P. & Zheng, N. Structure of DDB1 in complex with a Paramyxovirus V protein: viral hijack of a propeller cluster in ubiquitin ligase. *Cell* **124**, 105–117 (2006).
18. Scrima, A. *et al.* Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. *Cell* **135**, 1213–1223 (2008).
19. Li, T., Robert, E. I., van Breugel, P. C., Strubin, M. & Zheng, N. A promiscuous alpha-helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. *Nature Struct. Mol. Biol.* **17**, 105–111 (2010).
20. Scrima, A. *et al.* Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! *FEBS Lett.* **585**, 2818–2825 (2011).
21. Hur, S., Stroud, R. M. & Finer-Moore, J. Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: a structural perspective. *J. Biol. Chem.* **281**, 38969–38973 (2006).
22. Ruchelman, A. L. *et al.* Isosteric analogs of lenalidomide and pomalidomide: Synthesis and biological activity. *Bioorg. Med. Chem. Lett.* **23**, 360–365 (2013).
23. Jin, J., Arias, E., Chen, J., Harper, J. & Walter, J. A family of diverse Cul4–Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. *Mol. Cell* **23**, 709–721 (2006).
24. Angers, S. *et al.* Molecular architecture and assembly of the DDB1–CUL4A ubiquitin ligase machinery. *Nature* **443**, 590–593 (2006).
25. Fischer, E. S. *et al.* The molecular basis of CRL4DDB2/CSA ubiquitin ligase architecture, targeting, and activation. *Cell* **147**, 1024–1039 (2011).
26. Hornbeck, P. V. *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* **40**, D261–D270 (2012).
27. Hofmeister, C. C. *et al.* Phase I trial of lenalidomide and CCI-779 in patients with relapsed multiple myeloma: evidence for lenalidomide–CCI-779 interaction via P-glycoprotein. *J. Clin. Oncol.* **29**, 3427–3434 (2011).
28. Zhou, S. *et al.* Transport of thalidomide by the human intestinal caco-2 monolayers. *Eur. J. Drug Metab. Pharmacokinet.* **30**, 49–61 (2005).
29. Roche, S. *et al.* Development, validation and application of a sensitive LC-MS/MS method for the quantification of thalidomide in human serum, cells and cell culture medium. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **902**, 16–26 (2012).
30. Crowley, M. A. *et al.* Further evidence for the possible role of MEIS2 in the development of cleft palate and cardiac septum. *Am. J. Med. Genet. A* **152A**, 1326–1327 (2010).
31. Capdevila, J., Tsukui, T., Rodríguez Esteban, C., Zappavigna, V. & Izpisua Belmonte, J. C. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, 839–849 (1999).
32. Paige, S. L. *et al.* A temporal chromatin signature in human embryonic stem cells identifies regulators of cardiac development. *Cell* **151**, 221–232 (2012).
33. Skaar, J. R., Pagan, J. K. & Pagano, M. Mechanisms and function of substrate recruitment by F-box proteins. *Nature Rev. Mol. Cell Biol.* **14**, 369–381 (2013).
34. Bennett, E. J., Rush, J., Gygi, S. P. & Harper, J. W. Dynamics of Cullin–RING Ubiquitin Ligase Network Revealed by Systematic Quantitative Proteomics. *Cell* **143**, 951–965 (2010).
35. Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446**, 640–645 (2007).
36. Huai, Q. *et al.* Crystal structure of calcineurin–cyclophilin–cyclosporin shows common but distinct recognition of immunophilin–drug complexes. *Proc. Natl Acad. Sci. USA* **99**, 12037–12042 (2002).
37. Kissinger, C. R. *et al.* Crystal structures of human calcineurin and the human FKBP12–FK506–calcineurin complex. *Nature* **378**, 641–644 (1995).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Novartis Research Foundation and grants to N.H.T. from the European Research Council (ERC-2010-StG 260481–MoBa-CS) and to J.W.H. from the National Institutes of Health (AG011085). J.R.L. was supported by a Damon Runyon Postdoctoral Fellowship (DRG 2061–10). We acknowledge J. Reilly for performing immobilized artificial membrane (IAM) experiments. We thank J. Tallarico, J. Porter, W. Sellers, S. Cottens and M. Renatus for help and comments. D. Hess, R. Sack and J. Seebacher for the mass spectrometry analysis and J. Keusch and H. Gut for support. We thank W. Kaelin for kindly providing the *IKZF1* reporter plasmid (pCMV–IRES–*Renilla* luciferase–IRES–Gateway–firefly luciferase). Part of this work was performed at the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland.

Author Contributions E.S.F., N.H.T., J.L.J. and W.C.F. initiated the project. E.S.F. and K.B. conducted the protein purification and crystallization. G.M.L. provided recombinant CSN, and S.C. pre-screened protein complexes by electron microscopy. E.S.F. collected data and processed and refined X-ray data. E.S.F. and N.H.T. analysed the structures. E.S.F. performed *in vitro* experiments and, with the help of U.H., developed and performed TR-FRET and fluorescence polarization assays. E.S.F. performed protein array experiments. M.B.S. and E.S.F. analysed the data. E.S.F., K.B., J.R.L., H.Y., M.H., J.W.H. and N.H.T. conceived and performed the cell-biological characterization. R.B.T. and R.E.J.B. conceived and conducted the chemical syntheses. J.N. and M. Schirle performed proteomics. V.A. and J.O. carried out the differential scanning fluorimetry experiments. F.S. and M. Schebesta carried out the zebrafish experiments. E.S.F. and N.H.T. wrote the manuscript. All authors assisted in editing the manuscript.

Author Information Structural coordinates for human DDB1–G. *gallus* CRBN–thalidomide, human DDB1–G. *gallus* CRBN–lenalidomide and human DDB1–G. *gallus* CRBN–pomalidomide have been deposited in the Protein Data Bank under accession numbers 4CI1, 4CI2 and 4CI3. Human protein microarray data sets for this study have been deposited in the Gene Expression Omnibus database under accession number GSE57554. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.H.T. (nicolas.thoma@fmi.ch).

A luminous, blue progenitor system for the type Iax supernova 2012Z

Curtis McCully¹, Saurabh W. Jha¹, Ryan J. Foley^{2,3}, Lars Bildsten^{4,5}, Wen-fai Fong⁶, Robert P. Kirshner⁶, G. H. Marion^{6,7}, Adam G. Riess^{8,9} & Maximilian D. Stritzinger¹⁰

Type Iax supernovae are stellar explosions that are spectroscopically similar to some type Ia supernovae at the time of maximum light emission, except with lower ejecta velocities^{1,2}. They are also distinguished by lower luminosities. At late times, their spectroscopic properties diverge from those of other supernovae^{3–6}, but their composition (dominated by iron-group and intermediate-mass elements^{1,7}) suggests a physical connection to normal type Ia supernovae. Supernovae of type Iax are not rare; they occur at a rate between 5 and 30 per cent of the normal type Ia rate¹. The leading models for type Iax supernovae are thermonuclear explosions of accreting carbon–oxygen white dwarfs that do not completely unbind the star^{8–10}, implying that they are ‘less successful’ versions of normal type Ia supernovae, where complete stellar disruption is observed. Here we report the detection of the luminous, blue progenitor system of the type Iax SN 2012Z in deep pre-explosion imaging. The progenitor system’s luminosity, colours, environment and similarity to the progenitor of the Galactic helium nova V445 Puppi^{11–13} suggest that SN 2012Z was the explosion of a white dwarf accreting material from a helium-star companion. Observations over the next few years, after SN 2012Z has faded, will either confirm this hypothesis or perhaps show that this supernova was actually the explosive death of a massive star^{14,15}.

SN 2012Z was discovered¹⁶ in the Lick Observatory Supernova Search on 2012 January 29.15 UT. It had an optical spectrum similar to the type Iax (previously called SN 2002cx-like) SN 2005hk^{3–5} (see Extended Data Fig. 1). The similarities between type Iax and normal type Ia supernovae make understanding the progenitors of the former important, especially because no progenitor of the latter has been identified. Like core-collapse supernovae (but also slowly declining, luminous type Ia supernovae), type Iax supernovae are found preferentially in young, star-forming galaxies^{17,18}. A single type Iax supernova, SN 2008ge, was in a relatively old (S0) galaxy with no indication of current star formation to deep limits¹⁹. Non-detection of the progenitor of SN 2008ge in Hubble Space Telescope (HST) pre-explosion imaging restricts its initial mass to $\lesssim 12 M_{\odot}$ (where M_{\odot} is the solar mass), and combined with the lack of hydrogen or helium in the SN 2008ge spectrum, favours a white dwarf progenitor¹⁹.

Deep observations of NGC 1309, the host galaxy of SN 2012Z, were obtained with HST in 2005–06 and 2010, serendipitously including the location of the supernova before its explosion. To pinpoint the position of SN 2012Z with high precision, we obtained follow-up HST data in 2013. Colour-composite images made from these observations before and after the supernova are shown in Fig. 1, and photometry of stellar

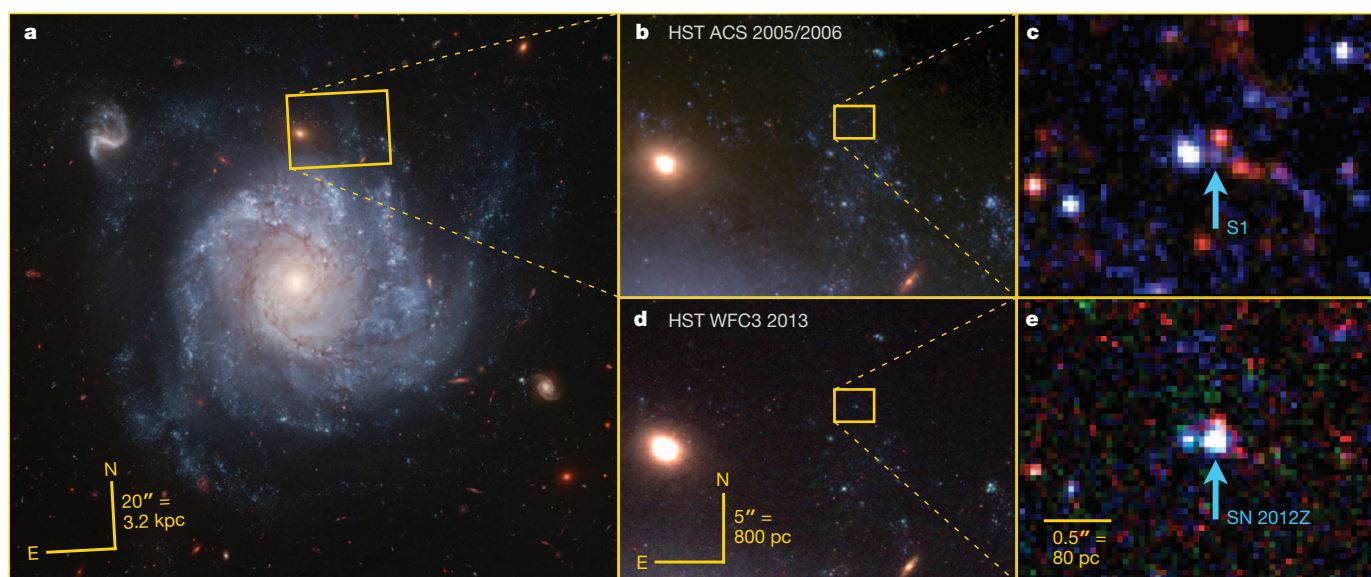


Figure 1 | HST colour images before and after supernova 2012Z. **a**, Hubble Heritage image of NGC 1309 (<http://heritage.stsci.edu/2006/07>); panels **b** and **c** zoom in on the progenitor system S1 in the deep, pre-explosion data.

d, e, Shallower post-explosion images of SN 2012Z on the same scale as **b** and **c**, respectively. The source data for these images are available as Supplementary Information.

¹Department of Physics and Astronomy, Rutgers, the State University of New Jersey, 136 Frelinghuysen Road, Piscataway, New Jersey 08854, USA. ²Astronomy Department, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, Illinois 61801, USA. ³Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, USA. ⁴Department of Physics, University of California, Santa Barbara, California 93106, USA. ⁵Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA. ⁶Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ⁷Department of Astronomy, University of Texas at Austin, Austin, Texas 78712, USA. ⁸Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. ⁹Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ¹⁰Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark.

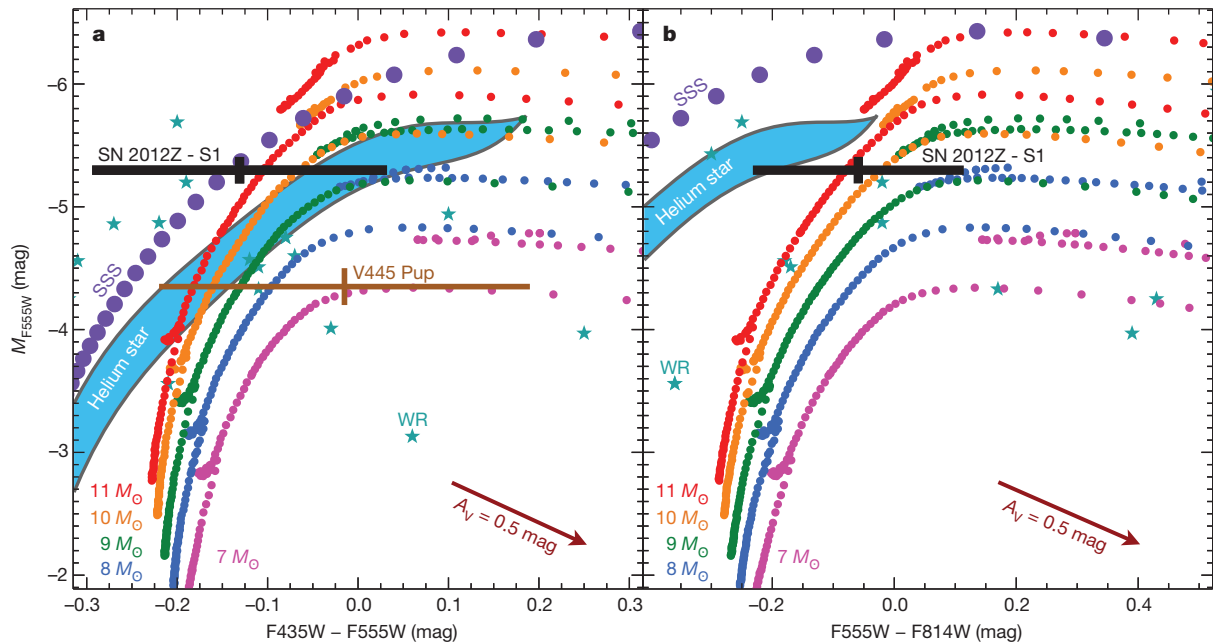


Figure 2 | Colour-magnitude diagrams of the SN 2012Z progenitor S1 and comparison models. **a**, The $F435W - F555W$ colour (roughly $B - V$), **b**, the $F555W - F814W$ colour (roughly $V - I$), both plotted against the $F555W$ (V) absolute magnitude. Black and brown crosses represent the progenitor systems for SN 2012Z and V445 Pup¹², respectively, with 1σ photometric uncertainties. Other comparisons plotted include evolutionary

tracks²⁴ for single stars (coloured dotted curves, with initial mass ranging from 7 to $11 M_{\odot}$ as indicated), thermal models for Eddington-luminosity super-soft sources (SSS; purple dots), candidate Wolf-Rayet stars²⁶ (WR; blue-grey stars), and models for helium-star donors to $1.2 M_{\odot}$ initial mass carbon-oxygen white dwarfs²⁹ (shaded blue regions). The effect of interstellar extinction with $A_V = 0.5$ mag is also shown (magenta arrow).

sources in the pre-explosion images near the supernova location is reported in Extended Data Table 1. We detect a source, called S1, coincident with the supernova at a formal separation of $0.0082'' \pm 0.0103''$ (equal to 1.3 ± 1.6 pc at 33 Mpc, the distance to NGC 1309; refs 20, 21). The pre-explosion data reach a 3σ limiting magnitude of $M_V \approx -3.5$ mag, quite deep for typical extragalactic supernova progenitor searches²², but certainly the possibility exists that the progenitor system of SN 2012Z was of lower luminosity and would be undetected in our data (as has been the case for all normal type Ia supernova progenitor searches to date²³). However, the locations of SN 2012Z and S1 are identical to within 0.8σ , and we estimate only a 0.24% (2.1%) probability that a random position near SN 2012Z would be within 1σ (3σ) of any detected star, making a chance alignment unlikely (see Methods, and Extended Data Fig. 2). We also observe evidence for variability in S1 (plausible for a pre-supernova system; Extended Data Table 2), at a level exhibited by only 4% of objects of similar brightness. We thus conclude there is a high likelihood that S1 is the progenitor system of SN 2012Z.

The colour-magnitude diagram (CMD) presented in Fig. 2 shows S1 to be luminous and blue, yet in an odd place on the diagram for a star about to explode. If its light is dominated by a single star, S1 is moderately consistent with an $\sim 18.5 M_{\odot}$ main-sequence star²⁴, an $\sim 11 M_{\odot}$ blue supergiant early in its evolution off the main sequence, or perhaps an $\sim 7.5 M_{\odot}$ (initial mass) blue supergiant later in its evolution (with core helium-burning in a blue loop, where models are quite sensitive to metallicity and rotation²⁵). None of these stars are expected to explode in standard stellar evolution theory, particularly without any signature of hydrogen in the supernova²².

The SN 2012Z progenitor system S1 is in a similar region in the CMD to some Wolf-Rayet stars²⁶: these are highly evolved, massive stars that are expected to undergo core collapse and may produce a supernova. If S1 were a single Wolf-Rayet star, its photometry is most consistent with the WN subtype and an initial mass of ~ 30 – $40 M_{\odot}$; such Wolf-Rayet stars are thought perhaps to explode with a helium-dominated outer layer as a type Ib supernova²⁷, and to be unlikely to produce the structure and composition of ejecta seen in type Iax supernovae^{1,3,6,7}. Moreover, isochrones²⁴ fitted to the neighbouring stars (Extended Data Fig. 3)

yield an age range of ~ 10 – 42 Myr, longer than the 5–8 Myr lifetime of such a massive Wolf-Rayet star.

S1 may be dominated by accretion luminosity; its brightness in B and V bands is not far from the predicted thermal emission of an Eddington-luminosity Chandrasekhar-mass white dwarf (a super-soft source, SSS in Fig. 2). However, its $V - I$ and $V - H$ colours are too red for a SSS model. A composite scenario, with accretion power dominating the blue flux, and another source providing the redder light (perhaps a fainter, red donor star) may be plausible.

The leading models of type Iax supernovae^{8–10} are based on explosions of carbon-oxygen white dwarfs, so S1 may be the companion star to an accreting white dwarf. Although there are a variety of potential progenitor systems (including main-sequence and red-giant donors, which are inconsistent with S1 if they dominate the system's luminosity), in standard scenarios no companion star can have an initial mass greater than $\sim 7 M_{\odot}$; otherwise, there would not be enough time to form the primary carbon-oxygen white dwarf that explodes. Thus, the photometry of S1 suggests that if it is the companion to a carbon-oxygen white dwarf, recent binary mass transfer must have played a role in its evolution. One model for a luminous, blue companion star is a relatively massive ($\sim 2 M_{\odot}$ when observed) helium star^{11,28,29}, formed after binary mass transfer and a common envelope phase (for example, a close binary with initial masses of ~ 7 and $\sim 4 M_{\odot}$). Although the model parameter space has not been fully explored, the predicted region in the CMD for helium star donors in a binary system with a $1.2 M_{\odot}$ initial-mass accreting carbon-oxygen white dwarf²⁹ is shown in Fig. 2, and S1 is consistent with being in this region. The evolutionary timescale for such a model is also well matched to the ages of nearby stars (Extended Data Fig. 3).

SN 2012Z and the star S1 have an interesting analogue in our own Milky Way, namely, the helium nova V445 Puppis^{11–13}, thought to be the surface explosion of a near-Chandrasekhar-mass helium-accreting white dwarf. Though S1 is somewhat brighter than the pre-explosion observations of V445 Pup, their consistent colours, similar variability amplitude¹³, and the physical connection between V445 Pup and likely progenitors of type Iax supernovae^{8–10} is highly suggestive. Indeed, two

type Ia supernovae (though not SN 2012Z itself) have shown evidence for helium in the system^{1,17}. In this model, a low helium accretion rate could lead to a helium nova (like V445 Pup), whereas a higher mass-transfer rate could result in stable helium burning on the carbon-oxygen white dwarf, allowing it to grow in mass before the supernova. The accretion is expected to begin as the helium star starts to evolve and grow in radius; indeed, S1 photometry is consistent with the evolutionary track of a helium star with a mass (after losing its hydrogen envelope) of $\sim 2 M_{\odot}$, on its way to becoming a red giant¹¹.

Although the scenario of a helium-star donor to an exploding carbon-oxygen white dwarf is a promising model for the progenitor and supernova observations, we cannot yet rule out the possibility that S1 is a single star that itself exploded. Fortunately, by late 2015, SN 2012Z will have faded below the brightness of S1, and HST imaging will allow us to distinguish between these models. Our favoured interpretation of S1 as the companion star predicts that it will still be detected (though perhaps modified by the impact of its exploding neighbour, a reduction in accretion luminosity, or a cessation of variability). On the other hand, if S1 has completely disappeared, it will be a strong challenge to models of type Ia supernovae, and will perhaps blur the line between thermonuclear white-dwarf supernovae and massive-star core-collapse supernovae, with important impacts on our understanding of stellar evolution and chemical enrichment.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 February; accepted 19 June 2014.

- Foley, R. J. *et al.* Type Ia supernovae: a new class of stellar explosion. *Astrophys. J.* **767**, 57 (2013).
- Li, W. *et al.* SN 2002cx: the most peculiar known type Ia supernova. *Publ. Astron. Soc. Pacif.* **115**, 453–473 (2003).
- Jha, S. *et al.* Late-time spectroscopy of SN 2002cx: the prototype of a new subclass of type Ia supernovae. *Astron. J.* **132**, 189–196 (2006).
- Phillips, M. M. *et al.* The peculiar SN 2005hk: do some type Ia supernovae explode as deflagrations? *Publ. Astron. Soc. Pacif.* **119**, 360–387 (2007).
- Sahu, D. K. *et al.* The evolution of the peculiar type Ia supernova SN 2005hk over 400 days. *Astrophys. J.* **680**, 580–592 (2008).
- McCully, C. *et al.* Hubble Space Telescope and ground-based observations of the type Ia supernovae SN 2005hk and SN 2008A. *Astrophys. J.* **786**, 134 (2014).
- Stritzinger, M. D. *et al.* Optical and near-IR observations of the faint and fast 2008ha-like supernova 2010ae. *Astron. Astrophys.* **561**, A146 (2014).
- Jordan, G. C., I. V., Perets, H. B., Fisher, R. T. & van Rossum, D. R. Failed-detonation supernovae: subluminal low-velocity Ia supernovae and their kicked remnant white dwarfs with iron-rich cores. *Astrophys. J.* **761**, L23 (2012).
- Kromer, M. *et al.* 3D deflagration simulations leaving bound remnants: a model for 2002cx-like type Ia supernovae. *Mon. Not. R. Astron. Soc.* **429**, 2287–2297 (2013).
- Fink, M. *et al.* Three-dimensional pure deflagration models with nucleosynthesis and synthetic observables for type Ia supernovae. *Mon. Not. R. Astron. Soc.* **438**, 1762–1783 (2014).
- Kato, M., Hachisu, I., Kiyota, S. & Saio, H. Helium nova on a very massive white dwarf: a revised light-curve model of V445 Puppis (2000). *Astrophys. J.* **684**, 1366–1373 (2008).
- Woudt, P. A. *et al.* The expanding bipolar shell of the helium nova V445 Puppis. *Astrophys. J.* **706**, 738–746 (2009).
- Goranskij, V., Shugarov, S., Zharova, A., Kroll, P. & Barsukova, E. A. The progenitor and remnant of the helium nova V445 Puppis. *Peremennye Zvezdy* **30**, 4 (2010); available at <http://www.astronet.ru/db/varstars/msg/eid/PZ-30-004>.
- Valenti, S. *et al.* A low-energy core-collapse supernova without a hydrogen envelope. *Nature* **459**, 674–677 (2009).
- Moriya, T. *et al.* fallback supernovae: a possible origin of peculiar supernovae with extremely low explosion energies. *Astrophys. J.* **719**, 1445–1453 (2010).
- Cenko, S. B. *et al.* Supernova 2012Z in NGC 1309 = Psn J03220535–1523156. *Cent. Bur. Electron. Electr.* **3014**, 1 (2012).
- Foley, R. J. *et al.* SN 2008ha: an extremely low luminosity and exceptionally low energy supernova. *Astron. J.* **138**, 376–391 (2009).
- Lyman, J. D. *et al.* Environment-derived constraints on the progenitors of low-luminosity type I supernovae. *Mon. Not. R. Astron. Soc.* **434**, 527–541 (2013).
- Foley, R. J. *et al.* On the progenitor and supernova of the SN 2002cx-like supernova 2008ge. *Astron. J.* **140**, 1321–1328 (2010).
- Riess, A. G. *et al.* Cepheid calibrations of modern type Ia supernovae: implications for the Hubble constant. *Astrophys. J. Suppl. Ser.* **183**, 109–141 (2009).
- Riess, A. G. *et al.* A 3% solution: determination of the Hubble constant with the Hubble Space Telescope and Wide Field Camera 3. *Astrophys. J.* **730**, 119 (2011).
- Smartt, S. J. Progenitors of core-collapse supernovae. *Annu. Rev. Astron. Astrophys.* **47**, 63–106 (2009).
- Li, W. *et al.* Exclusion of a luminous red giant as a companion star to the progenitor of supernova SN 2011fe. *Nature* **480**, 348–350 (2011).
- Bertelli, G., Nasi, E., Girardi, L. & Marigo, P. Scaled solar tracks and isochrones in a large region of the Z-Y plane. II. From 2.5 to 20 M_{\odot} stars. *Astron. Astrophys.* **508**, 355–369 (2009).
- Georgy, C. *et al.* Populations of rotating stars. I. Models from 1.7 to 15 M_{\odot} at $Z = 0.014$, 0.006, and 0.002 with $\Omega/\Omega_{\text{crit}}$ between 0 and 1. *Astron. Astrophys.* **553**, A24 (2013).
- Shara, M. M. *et al.* The vast population of Wolf-Rayet and red supergiant stars in M101. I. Motivation and first results. *Astron. J.* **146**, 162 (2013).
- Groh, J. H., Meynet, G., Georgy, C. & Ekström, S. Fundamental properties of core-collapse supernova and GRB progenitors: predicting the look of massive stars before death. *Astron. Astrophys.* **558**, A131 (2013).
- Iben, I. Jr & Tutukov, A. V. Helium star cataclysmics. *Astrophys. J.* **370**, 615–629 (1991).
- Liu, W.-M., Chen, W.-C., Wang, B. & Han, Z. W. Helium-star evolutionary channel to super-Chandrasekhar mass type Ia supernovae. *Astron. Astrophys.* **523**, A3 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We dedicate this paper to the memory of our friend and colleague Weidong Li, whose work on the Lick Observatory Supernova Search, SN 2002cx-like supernovae, and Hubble Space Telescope observations of supernova progenitors continues to inspire us. We thank the SH₂E team for assistance with data from HST programme GO-12880, E. Bertin for the development of the STIFF software to produce colour images, and A. Dolphin for software and guidance in photometry. This research at Rutgers University was supported through NASA/HST grant GO-12913.01 and National Science Foundation (NSF) CAREER award AST-0847157 to S.W.J.; NASA/HST grant GO-12999.01 to R.J.F. supported this work at the University of Illinois. At UC Santa Barbara, this work was supported by NSF grants PHY 11-25915 and AST 11-09174 to L.B. The Danish Agency for Science, Technology, and Innovation supported M.D.S. through a Sapere Aude Level 2 grant. Support for HST programmes GO-12913 and GO-12999 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

Author Contributions C.M., S.W.J. and R.J.F. performed the data analysis and were chiefly responsible for preparing the manuscript and figures. W.F., R.P.K., G.H.M. and A.G.R. assisted in developing the proposal to obtain HST observations, including acquiring supporting ground-based photometry and spectroscopy. L.B. provided insight into models for progenitor systems. M.D.S. analysed ground-based photometry and spectroscopy of the supernova, used as input for this paper. All authors contributed to extensive discussions about, and edits to, the paper draft.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.W.J. (saurabh@physics.rutgers.edu).

Tunable spin–spin interactions and entanglement of ions in separate potential wells

A. C. Wilson¹, Y. Colombe¹, K. R. Brown², E. Knill¹, D. Leibfried¹ & D. J. Wineland¹

Quantum simulation^{1,2}—the use of one quantum system to simulate a less controllable one—may provide an understanding of the many quantum systems which cannot be modelled using classical computers. Considerable progress in control and manipulation has been achieved for various quantum systems^{3–5}, but one of the remaining challenges is the implementation of scalable devices. In this regard, individual ions trapped in separate tunable potential wells are promising^{6–8}. Here we implement the basic features of this approach and demonstrate deterministic tuning of the Coulomb interaction between two ions, independently controlling their local wells. The scheme is suitable for emulating a range of spin–spin interactions, but to characterize the performance of our set-up we select one that entangles the internal states of the two ions with a fidelity of 0.82(1) (the digit in parentheses shows the standard error of the mean). Extension of this building block to a two-dimensional network, which is possible using ion-trap microfabrication processes⁹, may provide a new quantum simulator architecture with broad flexibility in designing and scaling the arrangement of ions and their mutual interactions. To perform useful quantum simulations, including those of condensed-matter phenomena such as the fractional quantum Hall effect, an array of tens of ions might be sufficient^{4,10,11}.

The use of effective spin–spin interactions between ions in separate potential wells is a key feature of proposals for simulation with two-dimensional systems of quantum spins with arbitrary conformations and versatile couplings^{6,7,12}. In addition, these effective spin–spin interactions may enable logic operations to be performed in a multi-zone quantum information processor^{13–15} without the need to bring the quantum bits (qubits) into the same trapping potential well^{16,17}. Such coupling might also prove useful for metrology and sensing. For example, it could extend the capabilities of quantum-logic spectroscopy^{18,19} to ions that cannot be trapped within the same potential as the measurement ion, such as oppositely charged ions or even antimatter particles¹⁸. Coupling could be obtained either through mutually shared electrodes^{18,20} or directly through the Coulomb interaction^{13,21}.

In the experiments described here, two ions of mass m are trapped at equilibrium distance d_0 in independent, approximately harmonic potential wells. Coulomb interaction between the ions leads to dipole–dipole-type coupling, with strength $\Omega_{\text{ex}} \propto d_0^{-3}$ (Methods), where the oscillations of the ions in their respective wells manifest the dipoles¹². The coupled system has six normal modes, four perpendicular to the direction between the double wells (radial) and two along this direction (axial). Although all these modes are useful for dipole–dipole coupling¹², we concentrate on the two axial modes, with uncoupled well frequencies $\omega_l \approx \omega_r$, and with eigenfrequencies and eigenvectors

$$\begin{aligned}\omega_{\text{str}} &= \bar{\omega} + \sqrt{\delta^2 + \Omega_{\text{ex}}^2} \\ \omega_{\text{com}} &= \bar{\omega} - \sqrt{\delta^2 + \Omega_{\text{ex}}^2} \\ q_{\text{str}} &= (\sin(\theta_{\text{str}}), \cos(\theta_{\text{str}})) \\ q_{\text{com}} &= (\sin(\theta_{\text{com}}), \cos(\theta_{\text{com}}))\end{aligned}\quad (1)$$

where $\theta_{\text{str}} = \arctan\left[\left(\delta - \sqrt{\delta^2 + \Omega_{\text{ex}}^2}\right)/\Omega_{\text{ex}}\right]$ and $\theta_{\text{com}} = \arctan\left[(\delta + \sqrt{\delta^2 + \Omega_{\text{ex}}^2})/\Omega_{\text{ex}}\right]$. The average well frequency is denoted by $\bar{\omega} \equiv (\omega_l + \omega_r)/2$, and the frequency difference is $2\delta = (\omega_r - \omega_l)$. For $|\delta| \gg \Omega_{\text{ex}}$ these modes decouple and the two ions move nearly independently of each other. When approaching resonance ($\delta = 0$), the motions of the ions are strongly coupled, resulting in an avoided crossing of the motional frequencies with a splitting of $2\Omega_{\text{ex}}$. On resonance, the normal modes are a centre-of-mass mode (ω_{com} , $q_{\text{com}} = (1/\sqrt{2}, 1/\sqrt{2})$) and a stretch mode (ω_{str} , $q_{\text{str}} = (-1/\sqrt{2}, 1/\sqrt{2})$), with motional quanta shared between the two ions.

These shared quantized degrees of freedom can simulate spin–spin interactions^{5,22,23}, just as for two-qubit quantum logic gates with ions in the same harmonic well; but, unlike in the latter case, the strength of the spin–spin interaction can be tuned from strong to weak by controlling the individual trapping wells^{12,16,17}. We denote the energy eigenstates of the pseudo-spin-1/2 systems as $|\uparrow\rangle, |\downarrow\rangle$, corresponding to internal states of the ions, separated by $\hbar\omega_0$ (\hbar , Planck's constant divided by 2π), and the number states of the normal modes as $|n_{\text{str}}\rangle$ and $|n_{\text{com}}\rangle$. We excite ‘carrier’ transitions $|\downarrow, n_{\text{str}}, n_{\text{com}}\rangle \leftrightarrow |\uparrow, n_{\text{str}}, n_{\text{com}}\rangle$ with a uniform oscillating field at the $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition frequency ω_0 , and with phase ϕ_c . Simultaneously, a single ‘red-sideband’ excitation at frequency $\omega_0 - \bar{\omega}$ and phase ϕ_s , between the sideband frequencies for the stretch and centre-of-mass modes, excites both the $|\downarrow, n_{\text{str}}, n_{\text{com}}\rangle \leftrightarrow |\uparrow, n_{\text{str}} - 1, n_{\text{com}}\rangle$ transition and the $|\downarrow, n_{\text{str}}, n_{\text{com}}\rangle \leftrightarrow |\uparrow, n_{\text{str}}, n_{\text{com}} - 1\rangle$ transition^{24,25}. These excitations emulate an effective spin–spin interaction (Methods)

$$\begin{aligned}\hat{H}_{\text{eff}} &= \hbar\kappa\hat{\sigma}_l^{\phi_c}\hat{\sigma}_r^{\phi_c} \\ \hat{\sigma}_{l/r}^{\phi_c} &= \cos(\phi_c)\hat{\sigma}_{l/r}^x - \sin(\phi_c)\hat{\sigma}_{l/r}^y\end{aligned}$$

where κ is the coupling strength and $\hat{\sigma}_{l/r}^{x/y}$ are the Pauli spin-1/2 operators of the respective ions. We can emulate antiferromagnetic ($\kappa > 0$) and ferromagnetic ($\kappa < 0$) interactions by our choice of the ion spacing or the detunings δ_{str} and δ_{com} of the normal modes relative to the sideband drive (Methods). Under the simultaneous carrier and red-sideband drive, the spins become periodically entangled and disentangled with the motion. Starting with a product state $|\Phi_i\rangle$, spins and motion are disentangled into a product state at $T_j = 2\pi j/\Omega_{\text{ex}}$ ($j > 0$ integer), but the spins acquire phases that depend on the ions’ motion in phase space during the off-resonance excitation. These phases simulate the spin–spin interaction²⁶. We benchmark our implementation of the spin–spin interaction by starting from the well-defined product state $|\Psi_i\rangle = |\downarrow\downarrow\rangle$, effectively evolving it under an antiferromagnetic ($\kappa > 0$) interaction for time $T_2 = \pi/4\kappa$ with $\phi_c = 0$, and comparing the resulting state with the maximally entangled state $|\Psi_e\rangle = \exp\left[-i\frac{\pi}{4}\hat{\sigma}_l^x\hat{\sigma}_r^x\right]|\Psi_i\rangle = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle - i|\uparrow\uparrow\rangle)$ that would be produced under ideal conditions (Methods).

The (pseudo-)spin-1/2 system is formed by the $|2s^2S_{1/2}, F=1, m_F=-1\rangle \equiv |\uparrow\rangle$ and $|2s^2S_{1/2}, F=2, m_F=-2\rangle \equiv |\downarrow\rangle$ hyperfine ground

¹National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305, USA. ²Georgia Tech Research Institute, 400 10th Street Northwest, Atlanta, Georgia 30332, USA.

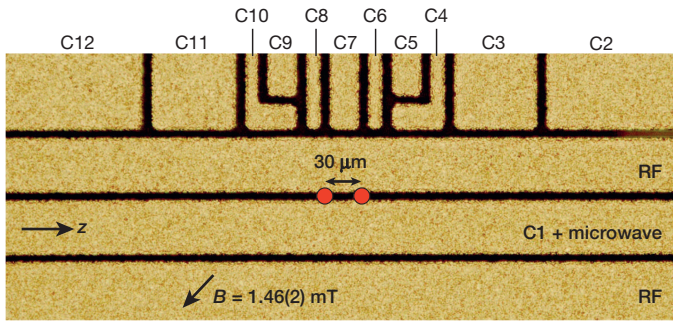


Figure 1 | Microfabricated surface-electrode trap. Microscope image of ion-trap electrodes, showing radio-frequency (RF) and static-potential control electrodes (C1–C12). Dark areas are the 5 μm gaps between electrodes. Ions are trapped 40 μm above the chip surface; red dots indicate the ion locations, with a 30 μm spacing. Electrode C1 also supports microwave currents at 1.28 GHz to drive carrier transitions on the two ions.

states of $^9\text{Be}^+$, where F is the total angular momentum and m_F is the component of F along a quantization axis provided by a 1.46(2) mT static magnetic field (Fig. 1). The ions are confined in a cryogenic (trap temperature < 5 K), microfabricated, surface-electrode linear Paul ion trap¹⁶ composed of 10 μm -thick gold electrodes separated by 5 μm gaps, deposited onto a crystalline quartz substrate. An oscillating potential (~ 100 V peak at 163 MHz), applied to the radiofrequency electrodes in Fig. 1, provides pseudopotential confinement of the ions in the radial (perpendicular to z) directions at motional frequencies of ~ 17 and ~ 27 MHz at a distance of approximately 40 μm from the trap surface. Along the trap z axis, a double well is formed by static potentials applied to control electrodes C1–C12. The axial (z) oscillation frequencies ω_1 and ω_r around the respective minima are typically near 4 MHz. Single-ion heating¹³ is in the range of 100 to 200 quanta per second. This heating is approximately four orders of magnitude larger than that due to our estimate of Johnson noise heating for this apparatus. For two ions spaced 30 μm apart, and in motional resonance ($\delta = 0$), the period required for the ions to exchange their motional energies is $\tau_{\text{ex}} = \pi/2\Omega_{\text{ex}} = 70$ μs , compared with an average period of 5–10 ms required to absorb a single motional quantum due to background heating. Fine adjustment of control-electrode potentials (at the 100 μV level) enables individual control of potential-well curvatures to tune the Coulomb interaction between the ions through resonance. Electrode C1 also supports microwave currents (typically of milliamperes amplitude) that produce an oscillating magnetic field to drive carrier transitions at the same rate in both ions.

Superimposed σ^- -polarized laser beams, nearly resonant with the $2s^2S_{1/2} \rightarrow 2p^2P_{1/2}$ and the $2s^2S_{1/2} \rightarrow 2p^2P_{3/2}$ transitions ($\lambda \approx 313$ nm) and propagating along the magnetic field direction, are used for optical pumping, Doppler laser cooling and state detection by resonance fluorescence. Optical pumping prepares both ions in $|\downarrow\rangle$. We can distinguish the $|\downarrow\rangle$ (bright) and $|\uparrow\rangle$ (dark) states by detecting resonance fluorescence on the $|\downarrow\rangle \rightarrow |2p^2P_{3/2}, F=3, m_F=-3\rangle$ optical cycling transition. Typically, three to five photons are detected per ion in $|\downarrow\rangle$ over a background of 0.15 to 0.6 photons on a photomultiplier during detection periods in the range 300–400 μs . A pair of elliptically shaped laser beams, separated in frequency by approximately the $|\downarrow\rangle \leftrightarrow |\uparrow\rangle$ transition frequency ($\omega_0 \approx 2\pi \times 1.28$ GHz) and detuned 80 GHz above the $^2S_{1/2} \rightarrow ^2P_{1/2}$ transition, illuminate both ions with equal intensity. These beams induce two-photon stimulated-Raman transitions for ground-state cooling¹³ and for the motional sideband excitations used to implement the spin-spin interaction²⁶. Derived from the same 313 nm source, the frequency difference between the beams is produced with acousto-optic modulators, and the beam orientation is such that the difference wavevector $k = k_2 - k_1$ is parallel to the z axis (with magnitude $k = 2\sqrt{2}\pi/\lambda$). The spin-spin coupling strength is $\kappa = \cos(2\phi)(\eta\Omega_s)^2/2\Omega_{\text{ex}}$, where $2\phi = kd_0$ is the phase difference of the beat note between the two laser fields at the positions of the ions, Ω_s is the stimulated-Raman Rabi frequency and $\eta = k/\sqrt{\hbar/2m\omega}$ (Methods).

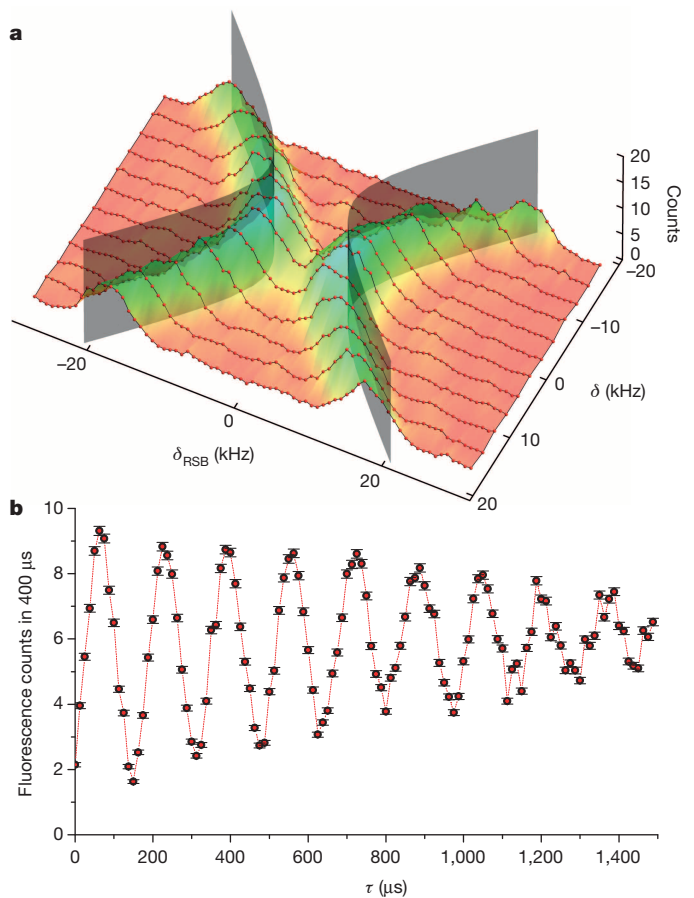


Figure 2 | Motional spectroscopy of two coupled ions. **a**, The red dots connected by black lines indicate separate scans of the red-sideband detuning δ_{RSB} from the average mode frequency $\bar{\omega}$ for different values of the difference δ between the individual well frequencies. The vertical scale is proportional to the sum of the probabilities for each ion to be in $|\downarrow\rangle$. At the centre of the avoided crossing, the normal mode frequency splitting Ω_{ex}/π is 12(1) kHz. Each data point represents an average of 200 experiments. Shaded planes are a theoretical prediction for the avoided crossing according to equations (1). **b**, Resonant ($\delta \approx 0$) single-quantum motional exchange between two ions, with an exchange time $\tau_{\text{ex}} = 80(2)$ μs . The vertical scale is proportional to the probability of the laser-addressed ion being in $|\downarrow\rangle$. Each data point represents an average of 500 experiments, and error bars correspond to s.e.m. Dashed lines are included to guide the eye.

A key to implementing spin-spin interactions with ions in separate trapping zones is being able to tune the well frequencies precisely enough to control the eigenfrequencies and eigenmodes (equations (1)) near the avoided crossing. In Fig. 2a, we characterize this avoided crossing. For these experiments, the ions are separated by 27(2) μm . They are laser-cooled nearly to their motional ground states (mean motional mode occupation, $\bar{n}_{\text{str/com}} \approx 0.1$), optically pumped to the $|\downarrow\downarrow\rangle$ state and then rotated into the $|\uparrow\uparrow\rangle$ state with a microwave carrier π -pulse. Fine adjustments are made to control electrodes C2 and C12 to tune the harmonic confinement of the two trapping zones, stepping the system through the avoided crossing. At each step, after cooling and optical pumping, we implement the Raman red-sideband drive and scan its detuning δ_{RSB} with respect to $\bar{\omega}$. If the sideband excitation frequency is equal to $\omega_0 - \omega_{\text{str}}$ or $\omega_0 - \omega_{\text{com}}$, then the spin of one or both ions can flip to $|\downarrow\rangle$ while absorbing quanta of motion, and a peak in the resonance fluorescence counts is observed. The spectral resolution is set by the duration of the square-pulse sideband excitation (120 μs). At the centre of the avoided crossing, the splitting of the mode frequencies is $2\Omega_{\text{ex}} = 2\pi \times 12(1)$ kHz.

In Fig. 2b, we show data that demonstrate single-phonon exchange between the two ions. With the trapping zones tuned to resonance ($\delta = 0$), both modes are cooled to near the motional ground state and the ions

are prepared in $|\uparrow\uparrow\rangle$. In this experiment, the two Raman beams are tightly focused onto only one of the ions and are used to add a single phonon to that ion (and flip its spin) with a π -pulse on the red sideband of its local frequency in a duration short compared with τ_{ex} . In this limit, after the pulse, the resulting motional state is an equal superposition of both modes, and the phonon energy is therefore exchanged between the ions with a period $2\tau_{\text{ex}}$ (ref. 16). To monitor the exchange, the same Raman interaction is applied again after a variable delay τ . This can flip the spin and remove the quantum of motion only if the motion resides solely in the addressed ion after a particular delay. The level of fluorescence is proportional to the probability of this spin flip. From this, we determine an exchange time of $\tau_{\text{ex}} = 80(2) \mu\text{s}$, consistent with an ion spacing of $30(2) \mu\text{m}$ for this experiment. The reduction in contrast for longer delays is caused mainly by fluctuations and drifts of the trapping potential. We estimate that $\delta/2\pi$ drifted by approximately 500 Hz (a significant fraction of $\Omega_{\text{ex}}/2\pi$) during the 2–3 minutes required for the 20,000 experiments that provided the data for Fig. 2b.

For benchmarking the spin–spin interaction, the laser beams for fluorescence detection, Doppler cooling and stimulated Raman transitions are made to spatially overlap both ions with equal intensity. The ion spacing (approximately $27 \mu\text{m}$ here) is adjusted to an integer number of half-wavelengths of the difference wavevector of the two Raman laser fields, by a technique described elsewhere²⁷, such that $\cos(2\phi) \approx 1$. The wells are tuned to resonance ($\delta = 0$) with adjustments to control electrodes C2 and C12. The ions are first Doppler-cooled, then Raman sideband-cooled to near the ground state on both normal modes, and finally optically pumped into the $|\downarrow\downarrow\rangle$ state. The spin–spin interaction is implemented by simultaneously applying a relatively strong resonant microwave carrier excitation (Rabi frequency, $\Omega_c = 2\pi \times 23.1(2) \text{ kHz}$) and an optical sideband excitation at $\omega_0 - \bar{\omega}$ (Rabi frequency, $\eta\Omega_s = 2\pi \times 2.4(2) \text{ kHz}$). The exchange frequency satisfies $2\Omega_{\text{ex}} = 2\pi \times 13(1) \text{ kHz}$, so that $\kappa = 2\pi \times 446(13) \text{ Hz}$. In the middle of the coupling period, we shift the phases ϕ_c and ϕ_s of both driving fields by 180° relative to their phases during the first half of the coupling period. These phase reversals suppress the dependence of the final state on the carrier Rabi frequency and reduce sensitivity of the spin–spin interaction to drifts in the detuning and the coupling time (Methods). At the end of the coupling period, fluorescence detection and subsequent fitting of the photon-count histograms to those for the three possible outcomes (two ions bright, $|\downarrow\downarrow\rangle$; one ion bright, $|\downarrow\uparrow\rangle$ or $|\uparrow\downarrow\rangle$; or both ions dark, $|\uparrow\uparrow\rangle$) yield the respective probabilities P_2 , P_1 and P_0 .

Evolution of these probabilities as functions of the coupling duration is shown in Fig. 3a. Near $300 \mu\text{s}$, P_2 and P_0 are approximately equal ($P_2 + P_0 = 0.91(2)$) and P_1 has reached a minimum. To show that the resulting state is entangled, in a subsequent experiment we stop the evolution at $300 \mu\text{s}$, apply a carrier $\pi/2$ -pulse of variable phase ϕ_a , and determine the parity $\Pi = P_2 + P_0 - P_1$ as a function of ϕ_a . These data are shown in Fig. 3b together with a fit to $A\cos(2\phi_a + \phi_0) + B$. The fitted probabilities and the contrast $A = 0.73(2)$ imply a state fidelity²⁸ $F = \langle \Psi_e | \rho_e | \Psi_e \rangle = (P_2 + P_0 + A)/2 = 0.82(1)$, where the density matrix ρ_e describes the experimentally produced state (Methods). From simulations and independent measurements, we estimate the leading contributions to the observed infidelity as follows: drift and fluctuations of the trapping potentials (including ‘anomalous’ motional heating) contribute ~ 0.08 ; spontaneous emission due to off-resonance excitation by Raman laser beams contributes ~ 0.02 ; Raman laser beam intensity fluctuations contribute ~ 0.03 ; and state preparation and detection errors contribute ~ 0.03 .

For scalable implementations of lattices of interacting spins, the quality and ease of tuning of the spin–spin interaction must be improved; however, there are no apparent fundamental barriers to this. Trap potential fluctuations in our experiments appear to be dominated by changes in surface charging and work functions rather than changes in externally applied control potentials. It should be possible to suppress these fluctuations by improving the surface quality of the electrodes²⁹, reducing the amount of nearby dielectric materials and minimizing the exposure

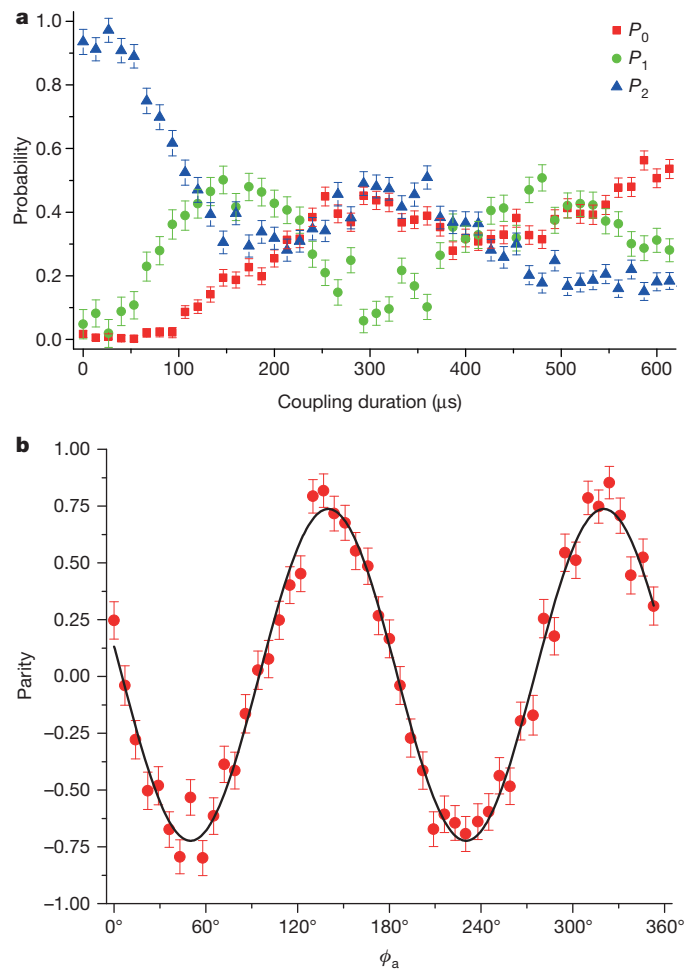


Figure 3 | Characterizing the spin–spin coupling interaction between ions in separate trapping zones. **a**, Evolution of probabilities P_0 of $|\uparrow\uparrow\rangle$ (red), P_1 of $|\downarrow\downarrow\rangle$ and $|\uparrow\downarrow\rangle$ (green), and P_2 of $|\downarrow\uparrow\rangle$ (blue), as functions of coupling duration. Each data point represents an average of 400 experiments, and error bars correspond to the s.e.m. **b**, Parity oscillation obtained (for a coupling duration of $300 \mu\text{s}$) by applying a carrier analysis $\pi/2$ -pulse with variable phase ϕ_a , and a fit to the data (black curve). Each data point represents an average of 400 experiments, and error bars correspond to the s.e.m.

of the electrodes to ultraviolet light through better beam shaping. Laser intensity and pointing noise can be reduced by passive or active stabilization of the beams with respect to the ions (or both), or potentially avoided entirely by using microwave gradient fields for the sideband interactions¹². The microfabrication techniques used to construct the trap are scalable to larger arrays of trapped ions, thus potentially enabling informative ‘analogue’ quantum simulations⁴ without requiring arbitrarily precise quantum control. Theoretical work to quantify the common belief that many observables of interest in analogue quantum simulations are sufficiently robust is ongoing³⁰ (Methods). Initial indications are that the proposed technical improvements may be sufficient. A three-by-three lattice is sufficient to simulate quantum Hall physics, and with six-by-six lattices fractional Hall effects and other intriguing solid-state phenomena become accessible^{8,11}. Even for these modest numbers of spins, modelling of quantum interactions with conventional computers is challenging; this difficulty may be overcome with quantum simulations.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper

Received 5 February; accepted 2 June 2014.

1. Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).

2. Lloyd, S. Universal quantum simulators. *Science* **273**, 1073–1078 (1996).
3. Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
4. Georgescu, I. M., Ashhab, S. & Nori, F. Quantum simulation. *Rev. Mod. Phys.* **86**, 153–185 (2014).
5. Blatt, R. & Roos, C. F. Quantum simulations with trapped ions. *Nature Phys.* **8**, 277–284 (2012).
6. Chiaverini, J. & Lybarger, W. E. Laserless trapped-ion quantum simulations without spontaneous scattering using microtrap arrays. *Phys. Rev. A* **77**, 022324 (2008).
7. Schmied, R., Wesenberg, J. H. & Leibfried, D. Optimal surface-electrode trap lattices for quantum simulation with trapped ions. *Phys. Rev. Lett.* **102**, 233002 (2009).
8. Shi, T. & Cirac, J. I. Topological phenomena in trapped-ion systems. *Phys. Rev. A* **87**, 013606 (2013).
9. Seidelin, S. *et al.* Microfabricated surface-electrode ion trap for scalable quantum information processing. *Phys. Rev. Lett.* **96**, 253003 (2006).
10. Friedenauer, A., Schmitz, H., Glueckert, J. T., Porras, D. & Schaetz, T. Simulating a quantum magnet with trapped ions. *Nature Phys.* **4**, 757–761 (2008).
11. Nielsen, A. E. B., Sierra, G. & Cirac, J. I. Local models of fractional quantum Hall states in lattices and physical implementation. *Nature Commun.* **4**, 2864 (2013).
12. Schmied, R., Wesenberg, J. H. & Leibfried, D. Quantum simulation of the hexagonal Kitaev model with trapped ions. *New J. Phys.* **13**, 115011 (2011).
13. Wineland, D. J. *et al.* Experimental issues in coherent quantum-state manipulation of trapped atomic ions. *J. Res. Natl Inst. Stand. Technol.* **103**, 259–328 (1998).
14. Cirac, J. I. & Zoller, P. A scalable quantum computer with ions in an array of microtraps. *Nature* **404**, 579–581 (2000).
15. Kielpinski, D., Monroe, C. & Wineland, D. J. Architecture for a large-scale ion-trap quantum computer. *Nature* **417**, 709–711 (2002).
16. Brown, K. R. *et al.* Coupled quantized mechanical oscillators. *Nature* **471**, 196–199 (2011).
17. Harlander, M., Lechner, R., Brownnutt, M., Blatt, R. & Hänsel, W. Trapped-ion antennae for the transmission of quantum information. *Nature* **471**, 200–203 (2011).
18. Heinzen, D. J. & Wineland, D. J. Quantum-limited cooling and detection of radio-frequency oscillations by laser-cooled ions. *Phys. Rev. A* **42**, 2977–2994 (1990).
19. Schmidt, P. O. *et al.* Spectroscopy using quantum logic. *Science* **309**, 749–752 (2005).
20. Daniilidis, N., Lee, T., Clark, R., Narayanan, S. & Häffner, H. Wiring up trapped ions to study aspects of quantum information. *J. Phys. B* **42**, 154012 (2009).
21. Ciaramicoli, G., Marzoli, I. & Tombesi, P. Scalable quantum processor with trapped electrons. *Phys. Rev. Lett.* **91**, 017901 (2003).
22. Kim, K. *et al.* Quantum simulation of frustrated Ising spins with trapped ions. *Nature* **465**, 590–593 (2010).
23. Britton, J. W. *et al.* Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature* **484**, 489–492 (2012).
24. Bermudez, A., Schmidt, P. O., Plenio, M. B. & Retzker, A. Robust trapped-ion quantum logic gates by continuous dynamical decoupling. *Phys. Rev. A* **85**, 040302(R) (2012).
25. Tan, T. R. *et al.* Demonstration of a dressed-state phase gate for trapped ions. *Phys. Rev. Lett.* **110**, 263002 (2013).
26. Porras, D. & Cirac, J. I. Effective quantum spin systems with trapped ions. *Phys. Rev. Lett.* **92**, 207901 (2004).
27. Chou, C. W., Hume, D. B., Thorpe, M. J., Wineland, D. J. & Rosenband, T. Quantum coherence between two atoms beyond $Q = 10^{15}$. *Phys. Rev. Lett.* **106**, 160801 (2011).
28. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
29. Hite, D. A. *et al.* 100-fold reduction of electric-field noise in an ion trap cleaned with in situ argon-ion-beam bombardment. *Phys. Rev. Lett.* **109**, 103001 (2012).
30. Hauke, P. *et al.* Can one trust quantum simulators? *Rep. Prog. Phys.* **75**, 082401 (2012).

Acknowledgements We thank K. McCormick, A. Keith and D. Allcock for comments on the manuscript. This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), ONR, and the NIST Quantum Information Program. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA or the ODNI. This work, a submission of NIST, is not subject to US copyright.

Author Contributions A.C.W. and D.L. designed the experiment, developed components of the experimental apparatus, collected data, analysed results and wrote the manuscript. D.L. developed the theory. Y.C. fabricated the ion-trap chip. K.R.B. built components of the apparatus, most notably the cryostat, and participated in the early design phase of the experiment. E.K. assisted with data analysis. D.J.W. participated in the design and analysis of the experiment. All authors discussed the results and the text of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C.W. (andrew.wilson@nist.gov).

METHODS

Normal modes of the coupled wells. We consider two ions, cooled close to their motional ground states. Along the direction of separation, each ion is confined to a separate minimum of a double-well potential with minima denoted 'l' (left) and 'r' (right). We assume much stronger confinement in the remaining directions, such that it is sufficient to consider only motion along the direction of the separated double well. The Hamiltonian of the motion of two ions of mass m and charge Q , spaced at an average distance d_0 in wells with local harmonic oscillator ladder operators \hat{a}_l and \hat{a}_r and uncoupled oscillation frequencies ω_l and ω_r , including Coulomb coupling and neglecting constant energy terms, can be written for small motional excitation as¹⁶

$$\hat{H}_m = \hbar\omega_l \hat{a}_l^\dagger \hat{a}_l + \hbar\omega_r \hat{a}_r^\dagger \hat{a}_r - \hbar\Omega_{\text{ex}} (\hat{a}_l^\dagger \hat{a}_r + \hat{a}_r^\dagger \hat{a}_l)$$

with

$$\Omega_{\text{ex}} = \frac{Q^2}{4\pi\epsilon_0 m \sqrt{\omega_l \omega_r} d_0^3}$$

We define $\bar{\omega} \equiv (\omega_l + \omega_r)/2$ and $\delta \equiv (\omega_r - \omega_l)/2$ and transform the motion into a normal-mode basis with eigenfrequencies and eigenvectors (expressed in the eigenmode basis of two uncoupled ions)

$$\omega_{\text{str/com}} = \bar{\omega} \pm \sqrt{\delta^2 + \Omega_{\text{ex}}^2}$$

$$q_{\text{str/com}} = (q_{\text{str/com}}^{(l)}, q_{\text{str/com}}^{(r)}) = (\sin(\theta_{\text{str/com}}), \cos(\theta_{\text{str/com}}))$$

where

$$\theta_{\text{str/com}} = \arctan \left[\frac{\delta \mp \sqrt{\delta^2 + \Omega_{\text{ex}}^2}}{\Omega_{\text{ex}}} \right]$$

and the upper and lower signs apply to the stretch and centre-of-mass modes, respectively. In this basis, the motional Hamiltonian is

$$\hat{H}_m = \hbar\omega_{\text{str}} \hat{a}_{\text{str}}^\dagger \hat{a}_{\text{str}} + \hbar\omega_{\text{com}} \hat{a}_{\text{com}}^\dagger \hat{a}_{\text{com}}$$

where $\hat{a}_{\text{str/com}}$ are the corresponding ladder operators in the coupled basis. For $\delta = 0$, we recover the familiar centre-of-mass and stretch modes with a mode splitting of $2\Omega_{\text{ex}}$ and in the limit $\delta \ll \Omega_{\text{ex}}$ we can approximate

$$q_{\text{str/com}} \approx \left(\frac{\mp 1}{\sqrt{2}} \left(1 \mp \frac{\delta}{2\Omega_{\text{ex}}} \right), \frac{1}{\sqrt{2}} \left(1 \pm \frac{\delta}{2\Omega_{\text{ex}}} \right) \right)$$

Interaction Hamiltonian. The two ions are driven resonantly by a spatially uniform excitation on the carrier transition $|\uparrow_{l/r}\rangle \leftrightarrow |\downarrow_{l/r}\rangle = \hat{\sigma}_{l/r}^\pm |\uparrow_{l/r}\rangle$ at frequency ω_0 , Rabi-frequency Ω_c and phase ϕ_c . In the interaction picture and rotating-wave approximation, the carrier interaction takes the form

$$\hat{H}_c = \hbar\Omega_c [(\hat{\sigma}_l^- + \hat{\sigma}_r^-) e^{-i\phi_c} + (\hat{\sigma}_l^+ + \hat{\sigma}_r^+) e^{i\phi_c}]$$

with $\hat{\sigma}_{l/r}^\pm = (\hat{\sigma}_{l/r}^\pm)^\dagger$. Simultaneously, the ions are driven close to the Raman red sidebands of both normal modes by two laser beams (quantities associated with which will be denoted using indices 1 and 2) with difference wavevector ($k = k_2 - k_1$; magnitude $k = 2\sqrt{2}\pi/\lambda$) aligned along the direction of the double well, having frequency difference $\Delta\omega_L = \omega_2 - \omega_1 \approx \omega_0 - \bar{\omega}$, and phase difference $2\phi = kd_0$ for the beat note between the two laser fields at the positions of the ions. For $\Delta\omega_L = \omega_0$, the carrier Rabi rate is Ω_s . We assume the Lamb–Dicke limit, where $(\eta_{\text{str/com}} q_{\text{str/com}}^{(l/r)})^2 \bar{n}_{\text{str/com}} \ll 1$, with $\bar{n}_{\text{str/com}}$ the average occupation numbers and $\eta_{\text{str/com}} = k\sqrt{\hbar/2m\omega_{\text{str/com}}}$, the Lamb–Dicke parameters of the respective normal modes. The near-resonant terms of the red-sideband Hamiltonian are

$$\begin{aligned} \hat{H}_{\text{rsb}} = & i\hbar\Omega_s \left[\eta_{\text{com}} q_{\text{com}}^{(l)} \hat{a}_{\text{com}} \hat{\sigma}_l^+ e^{-i(\delta_{\text{com}} t - \phi_s + \phi)} \right. \\ & + \eta_{\text{str}} q_{\text{str}}^{(l)} \hat{a}_{\text{str}} \hat{\sigma}_l^+ e^{-i(\delta_{\text{str}} t - \phi_s + \phi)} \\ & + \eta_{\text{com}} q_{\text{com}}^{(r)} \hat{a}_{\text{com}} \hat{\sigma}_r^+ e^{-i(\delta_{\text{com}} t - \phi_s - \phi)} \\ & \left. + \eta_{\text{str}} q_{\text{str}}^{(r)} \hat{a}_{\text{str}} \hat{\sigma}_r^+ e^{-i(\delta_{\text{str}} t - \phi_s - \phi)} \right] + \text{h.c.} \end{aligned}$$

where $\delta_{\text{str/com}} = \Delta\omega_L - \omega_0 + \omega_{\text{str/com}}$ is the detuning relative to the red sideband of the respective normal mode, and ϕ_s is the phase of the sideband excitation at the mean position of the ions.

Spin–spin interaction. In the limit of a strongly driven carrier, such that $|\Omega_c| \gg \left\{ \eta_{\text{str/com}} \sqrt{\bar{n}_{\text{str/com}}} \Omega_s, |\delta_{\text{str/com}}| \right\}$, it is helpful to first transform to an internal-state basis where the bare spin states are dressed by the carrier^{24,25}. In this dressed frame, the basis states $\{|\pm_{l/r}\rangle, |\pm_{l/r}\rangle\}$ are eigenstates of

$$\hat{\sigma}_{(l/r)}^\phi = \cos(\phi_c) \hat{\sigma}_{(l/r)}^x - \sin(\phi_c) \hat{\sigma}_{(l/r)}^y$$

with $|\pm_{l/r}\rangle = \frac{1}{\sqrt{2}} (|\uparrow_{l/r}\rangle \pm e^{-i\phi_c} |\downarrow_{l/r}\rangle)$ and $\hat{\sigma}_{(l/r)}^\phi |\pm_{l/r}\rangle = \pm |\pm_{l/r}\rangle$. For each of the four internal basis states $|\pm_{l/r}\rangle$ and each normal mode, the sideband interaction can be written (neglecting rapidly oscillating terms near $2\Omega_c$)

$$\begin{aligned} \hat{H}_d = & i\hbar (d_{\text{com}} e^{i\delta_{\text{com}} t} \hat{a}_{\text{com}}^\dagger - d_{\text{com}}^* e^{-i\delta_{\text{com}} t} \hat{a}_{\text{com}}) \\ & + i\hbar (d_{\text{str}} e^{i\delta_{\text{str}} t} \hat{a}_{\text{str}}^\dagger - d_{\text{str}}^* e^{-i\delta_{\text{str}} t} \hat{a}_{\text{str}}) \end{aligned}$$

where the coefficients $d_{\text{str/com}}$ are state-dependent coherent displacement rates

$$\begin{aligned} d_{\text{str/com}}(s_l, s_r) = & -\frac{\Omega_s}{2} \eta_{\text{str/com}} \left(\sin(\theta_{\text{str/com}}) s_l e^{-i(\phi_s - \phi_c - \phi)} \right. \\ & \left. + \cos(\theta_{\text{str/com}}) s_r e^{-i(\phi_s - \phi_c + \phi)} \right) \end{aligned} \quad (2)$$

with $s_{l/r} \in \{-1, 1\}$ the eigenvalues corresponding to the basis states in question. The integrated displacements $\alpha_{\text{str/com}}$ and the geometric phases $\Phi_{\text{str/com}}$ acquired after time t are³¹

$$\begin{aligned} \alpha_{\text{str/com}}(s_l, s_r, t) = & i \frac{d_{\text{str/com}}(s_l, s_r)}{\delta_{\text{str/com}}} (1 - e^{i\delta_{\text{str/com}} t}) \\ \Phi_{\text{str/com}}(s_l, s_r, t) = & \frac{|d_{\text{str/com}}(s_l, s_r)|^2}{\delta_{\text{str/com}}^2} (\delta_{\text{str/com}} t - \sin(\delta_{\text{str/com}} t)) \end{aligned} \quad (3)$$

To return the motions of both modes to the original state after an interaction duration T , we require $\alpha_{\text{str/com}}(s_l, s_r, T) = 0$. This happens irrespective of the (state-dependent) magnitude of $d_{\text{str/com}}$ if $\delta_{\text{str/com}} T = c_{\text{str/com}} (2\pi)$ with $c_{\text{str/com}}$ an integer. In such cases, the motion is displaced around $|c_{\text{str/com}}|$ full circles in the respective phase spaces of the two modes by the interaction. Also, because $\delta_{\text{str}} - \delta_{\text{com}} = 2\sqrt{\delta^2 + \Omega_{\text{ex}}^2}$, the interaction duration can assume only certain values, determined by $\Delta c \equiv c_{\text{str}} - c_{\text{com}} > 0$, for the motion to return to its original state:

$$T = \frac{\pi \Delta c}{\sqrt{\delta^2 + \Omega_{\text{ex}}^2}}$$

If the spin and motional states are in a product state initially, they will be in a product state at T and any integer multiple of T . The spin-dependent phases acquired during T simplify to

$$\begin{aligned} \Phi_{\text{str/com}}(s_l, s_r) = & \left(\frac{\eta_{\text{str/com}} \Omega_s}{2} \right)^2 \\ & \times \frac{T}{\delta_{\text{str/com}}} [1 + s_l s_r \cos(2\phi) \sin(2\theta_{\text{str/com}})] \end{aligned}$$

The spin-dependent term is largest if $\phi = j\pi/2$ with j integer. This corresponds to the ions being spaced by an integer number of half-wavelengths π/k . In the experiment, the separation of the ions is controlled by slight changes in the well curvatures to ensure half-integer wavelength spacing. Also, $|\sin(2\theta_{\text{str/com}})|$ is reduced for $|\delta| > 0$ and eventually vanishes as the modes decouple in the limit $|\delta| \gg \Omega_{\text{ex}}$; therefore, the most efficient spin–spin interactions are implemented for $\delta = 0$. For our experimental conditions and $\delta = 0$, the mode splitting is much smaller than the average mode frequency $\bar{\omega}$, so we can approximate $\eta_{\text{str/com}} \approx \eta = k\sqrt{\hbar/2m\bar{\omega}}$. If we also assume that $\delta \ll \Omega_{\text{ex}}$, the phases simplify to

$$\begin{aligned} \Phi_{\text{str/com}}(s_l, s_r, T) = & \left(\frac{\eta \Omega_s}{2} \right)^2 \frac{T}{\delta_{\text{str/com}}} \left[1 \mp s_l s_r \cos(2\phi) \left(1 - \frac{\delta^2}{2\Omega_{\text{ex}}^2} \right) \right] \end{aligned}$$

In this limit, the phases $\Phi_{\text{str/com}}(s_l, s_r, T)$ depend only to second order on the relative detuning of the two wells. The shortest loop duration T is realized for $\Delta c = 1$, but the phase accumulates most effectively when the sideband drive is tuned to $\bar{\omega}$, exactly halfway between the normal modes ($c_{\text{str/com}} = \pm 1$, $\Delta c = 2$). At this detuning, the logical phase acquired on both modes adds constructively, and there is always some degree of phase cancellation for all other possible settings of the detuning. The total phase accumulated on both modes during T is

$$\begin{aligned}\Phi(s_l, s_r, T) &= \Phi_{\text{str}}(s_l, s_r, T) + \Phi_{\text{com}}(s_l, s_r, T) \\ &= -\cos(2\phi) \frac{(\eta\Omega_s)^2}{2\Omega_{\text{ex}}} s_l s_r T\end{aligned}$$

For any integer multiple of T , we can summarize the action of the applied fields as

$$|\pm 1, \pm r, jT\rangle = \exp\left[-i\cos(2\phi) \frac{(\eta\Omega_s)^2}{2\Omega_{\text{ex}}} \hat{\sigma}_l^{\phi_c} \hat{\sigma}_r^{\phi_c} jT\right] |\pm 1, \pm r, 0\rangle$$

with j a positive integer. Because this holds for a complete set of spin-basis states, it also holds for any general initial state of the system. Therefore, at any multiple of T , the system evolution is equivalent to that under the spin-spin Hamiltonian

$$\hat{H}_{\text{eff}} = \hbar\kappa \hat{\sigma}_l^{\phi_c} \hat{\sigma}_r^{\phi_c} \quad (4)$$

$$\kappa = \cos(2\phi) \frac{(\eta\Omega_s)^2}{2\Omega_{\text{ex}}} \quad (5)$$

A change from ferromagnetic to anti-ferromagnetic interaction can be accomplished by a π/k change in the ion spacing, corresponding to a $\pi/2$ change in ϕ . Alternatively, for example, $\kappa' = -\kappa/3 < 0$ is realized with a choice of detuning such that ($c_{\text{str}} = -1$, $c_{\text{com}} = -3$).

In principle, we can either perform a ‘stroboscopic’ emulation with the total duration a multiple of T , or use detunings $\delta_{\text{str/com}}$ whose magnitudes are much larger, so that all $|\alpha_{\pm}| \ll 1$ for any given time. For all multiples of T , the motional states of the ions factor from the spin states, so if one only ‘looks’ stroboscopically at times jT , the system effectively appears as though only the spins have evolved according to equations (4) and (5), while the motion has returned to its original state, thus appearing to have been unaffected. For much larger magnitude detunings $\delta_{\text{str/com}}$ spin-motion entanglement, and, thus, the deviation of the simulated state from that under the ideal spin-spin interaction, is small for arbitrary durations of the interaction²⁶. The added robustness comes at the expense of a weaker spin-spin interaction, which has to be compensated for by higher drive power or longer simulation timescales. Finally, rather than suppressing the bosonic harmonic oscillator modes, we can include them as an integral part of the simulator and study collective spin-boson Hamiltonians, which have been recently shown to contain complex behaviour comparable to models with only spin-spin interactions³².

Experimental characterization. We benchmark the spin-spin Hamiltonian of equations (4) and (5) by using it to entangle the hyperfine states (pseudo-spins) of two ions starting from the initial state $|\downarrow\downarrow\rangle$. To gain isolation from small errors, we break the total spin-spin interaction into two loops in phase space with $\kappa = \pi/8$ for each loop. For the first loop, we can choose $\phi_c = 0$ and $\phi_s = 0$ so that the eigenstates in the dressed basis are those of $\hat{\sigma}_{l/r}^x$. After finishing the first loop, we change carrier and sideband phases to $\phi_c = \phi_s = \pi$. The change in carrier phase is such that at the end of the second loop, the rotating frame due to the carrier is re-aligned with the frame of the bare states. This is because rotations around the x axis of the Bloch sphere in the first loop are unwound by rotating around the $-x$ axis for the same duration in the second loop. In addition, the phase change in the sideband drive ensures that $d_{\text{str/com}}(s_l, s_r)$ of the first loop is followed by $-d_{\text{str/com}}(s_l, s_r)$ in the second loop. In total there are three sign changes in the displacement rate equation (2), the first from replacing $\hat{\sigma}_{l/r}^x$ by $\hat{\sigma}_{l/r}^x = -\hat{\sigma}_{l/r}^x$ and therefore $s_{l/r} \rightarrow -s_{l/r}$, the second due to $\phi_c = 0 \rightarrow \pi$ and the third due to $\phi_s = 0 \rightarrow \pi$, which multiply to change the sign of the displacement rate. As a consequence, the total displacement $\alpha_{\text{str/com}}(s_l, s_r, T)$ in the second loop (equation (3)) is equal and opposite to that in the first loop and the motional wavefunctions return to their original positions in phase space even if $\alpha_{\text{str/com}}(s_l, s_r, T) \neq 0$ due to small errors in the detunings $\delta_{\text{str/com}}$ or in loop duration, provided that those errors are constant over both loops³³. The phases $\Phi_{\text{str/com}}(s_l, s_r)$ depend only on $|d_{\text{str/com}}(s_l, s_r)|^2$, and the effective spin-spin evolution is therefore the same in both loops. With the sideband excitation tuned to $\bar{\omega}$, a single loop duration corresponds to $T_L = 2\pi/\Omega_{\text{ex}}$ for a total interaction duration of $2T_L$. Starting from the initial state $|\downarrow\downarrow\rangle$, we would ideally produce the maximally entangled state $|\Psi_e\rangle = \exp\left[-i\frac{\pi}{4}\hat{\sigma}_l^x\hat{\sigma}_r^x\right]|\Psi_i\rangle = \frac{1}{\sqrt{2}}(|\downarrow\downarrow\rangle + |\uparrow\uparrow\rangle)$, if the sideband Rabi frequency satisfies $\eta\Omega_s = \Omega_{\text{ex}}/2\sqrt{2}$.

Determination of probabilities from state-dependent fluorescence. During one detection period (300–400 μs) we typically detect between 0.15 and 0.6 counts if both ions are projected into $|\uparrow\rangle$, and 3 to 5 additional counts for each ion in state $|\downarrow\rangle$. For each experimental setting, we record count histograms for 200–500 experiments.

Consider a count histogram $h = (h(i))_i$, where $h(i)$ experiments yielded i counts and $N = \sum_i h(i)$ is the total number of recorded counts. We infer the probabilities P_b with $b = 0, 1, 2$ by applying probability estimators $w_b = (w_b(i))_i$ to h according to $P_b = \sum_i w_b(i)h(i)/N$. The probability estimators are determined from the recorded

photon counts for on-resonance microwave Ramsey experiments with two ions, where the phase ϕ of the second $\pi/2$ -pulse was varied. These experiments are performed before and after the experiments to be analysed. An ideal such Ramsey experiment satisfies

$$P_0(\phi) = \cos^4(\phi/2)$$

$$P_1(\phi) = \sin^2(\phi)/2$$

$$P_2(\phi) = \sin^4(\phi/2)$$

The histograms h_ϕ recorded at phase ϕ are sampled from the mixture $P_0q_0 + P_1q_1 + P_2q_2$, where the q_b are the count distributions for zero, one or two bright ions. From this model and the Ramsey data, we can determine w_b so that $\sum_i w_b(i)h_\phi(i)$ yields $P_b(\phi)$. We use a linear least-squares fit, regularizing it to minimize the anticipated variance when inferring P_b for the completely mixed state.

Given a probability estimator w and a recorded histogram h , we estimate the experimental variance of the inferred probability P according to $v = (\sum_i w(i)^2 h(i)/N - P^2)/(N-1)$. This variance determines the error bars in Fig. 3. For the fidelities and related quantities, the variation in the probability estimators due to the finite statistics of the Ramsey experiments contributes an error comparable to this variance. To determine the overall statistical error in the fidelities, we used non-parametric bootstrap resampling³⁴ on all contributing histograms with 100 bootstrap resamples to determine error bars for fidelities and contrasts.

The assumed model for the Ramsey experiments makes no assumptions about the shapes or relationships of the count distributions q_b . This was important because we found that the q_b exhibit clear deviations from Poissonian distributions. We also determined c_b , the mean number of counts according to q_b , and found that $c_2 - c_0$ exceeded $2(c_1 - c_0)$ by about 8% for all the Ramsey scans considered.

Several effects result in deviations from an ideal Ramsey experiment. We found that there is a phase offset of approximately 5° in the Ramsey scans. We shifted the phase accordingly before determining the probability estimators. This had a statistically negligible effect on inferred probabilities and fidelities. After adjusting for the phase shift, we found no signature of a mismatch between the model and the data. In addition to checking that the dependence of the histograms on the phase was as expected, we considered whether there are more than three count distributions contributing to the Ramsey scans. We found no signature of such an effect. Furthermore, all other histograms, including those used to determine fidelities, could be explained as arising from a mixture of the same three count distributions.

An important effect that need not be apparent from the data is state-preparation error. By simulating Ramsey experiments with state-preparation error and q_b as inferred from the data, we determined that such errors lead to systematic overestimates of fidelities that are well correlated with the state-preparation error. The simulations involved initial states that are mixtures of the basis states. Let ε ($\ll 1$) be the probability that the state in this mixture is not $|\downarrow\downarrow\rangle$. For the inferred fidelities, we estimate a systematic increase in fidelity of approximately 1.1ε . The quoted systematic errors are based on a pessimistic upper bound of 0.01 on ε . In inferring P_b for a single histogram (as required for the plots in Fig. 3), these biases are small compared with the statistical error and were therefore not included in the error bars. We assumed that pulse errors had a statistically small effect on inferred probabilities and fidelities.

Discussion on robustness of analogue simulations. Richard Feynman stated that, “with a suitable class of quantum machines you could imitate any quantum system, including the physical world”³¹. For arbitrarily precise quantum simulations, this requires scalable quantum computers that employ error correction, but realizing these computers has proven to be very difficult. An alternative that may circumvent the difficulties is to faithfully map the dynamics of the physical model of interest onto sufficiently controllable quantum systems. This is called ‘analogue quantum simulation’. Because the overall physical properties of interest are often determined by local observables, the expectation is that the full quantum state need not be arbitrarily precise for useful information to be obtained³⁵. For example, although the global many-body state of the simulator is sensitive to a local perturbation, the expectation values of intensive properties can be more robust³⁰. It is also noteworthy that many material properties are robust in the presence of naturally occurring imperfections. This suggests that a useful analogue quantum simulator might be significantly easier to construct than a quantum computer, even in the absence of sufficiently precise quantum gates or explicit quantum error-correction strategies needed for fault tolerance³⁶.

Although the robustness of analogue quantum simulations is frequently asserted, it is not a simple matter to quantify the effects of experimental imperfections on physical properties of interest. At present, there does not exist a perfect and rigorous way to assess the quality of the results that one can expect from an analogue quantum simulation³⁰. Nevertheless, one can seek models and conditions for which the effects of the quantum simulator’s imperfections are expected to be minor and well

understood. A number of experimental groups, across multiple platforms, are currently pursuing this strategy. An alternative is to seek validation of the results on small systems that can be classically verified before obtaining results on large systems realizing the same model. In addition, validation may come from consistent results on multiple independent simulator platforms. This can eliminate simulator artefacts, as has been suggested in ref. 37.

For many developers of quantum simulators, a common Hamiltonian for testing their setups is the transverse Ising model^{4,8,10,22,23,26}. Recently, a theoretical investigation into the influence of disorder on the fidelity of quantum simulations of the Ising model was performed³⁰. With relatively large spin chains, analogue quantum simulator results are predicted to be usefully robust to random variation in the coupling coefficient up to a few per cent. This high tolerance to coupling imperfections, relative to a comparable universal quantum computation, is achieved because the simulation required that only local observables, rather than the entire simulator state, be robust. Although this work does not account for other technical issues that often limit the performance of experiments, it is nonetheless a useful performance indicator. In relation to our work, it suggests that although further progress on reducing experimental imperfections is probably required, the future technical improvements we propose may be sufficient. It may also be possible to ensure that the experimental imperfections correspond to physically relevant effects in the model under consideration. For example, Lloyd suggested that, “decoherence and thermal effects in the quantum computer can be exploited to mimic decoherence and

thermal effects in the system to be simulated”³², as was recently demonstrated³⁸. To ensure that the platform’s imperfections represent physically relevant interactions between the model and its normal environment, one can sometimes engineer the mapping from the ideal model to the experimental platform³⁹. Although we cannot make a general statement on the robustness of analogue quantum simulations, the above discussion is suggestive and many promising examples have been proposed.

31. Leibfried, D. *et al.* Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
32. Jünemann, J., Cadarso, A., Pérez-García, D., Bermudez, A. & García-Ripoll, J. J. Lieb-Robinson bounds for spin-boson lattice models and trapped ions. *Phys. Rev. Lett.* **111**, 230404 (2013).
33. Hayes, D. *et al.* Coherent error suppression in multiqubit entangling gates. *Phys. Rev. Lett.* **109**, 020503 (2012).
34. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, 1993).
35. Cirac, J. I. & Zoller, P. Goals and opportunities in quantum simulation. *Nature Phys.* **8**, 264–266 (2012).
36. Knill, E. Quantum computing. *Nature* **463**, 441–443 (2010).
37. Leibfried, D. Could a boom in technologies trap Feynman’s simulator? *Nature* **463**, 608 (2010).
38. Li, J. *et al.* Motional averaging in a superconducting qubit. *Nature Commun.* **4**, 1420 (2013).
39. Tseng, C. H. *et al.* Quantum simulation with natural decoherence. *Phys. Rev. A* **62**, 032309 (2000).

Controlled synthesis of single-chirality carbon nanotubes

Juan Ramon Sanchez-Valencia^{1†}, Thomas Dienel¹, Oliver Gröning¹, Ivan Shorubalko², Andreas Mueller^{3†}, Martin Jansen³, Konstantin Amsharov^{3†}, Pascal Ruffieux¹ & Roman Fasel^{1,4}

Over the past two decades, single-walled carbon nanotubes (SWCNTs) have received much attention because their extraordinary properties are promising for numerous applications^{1,2}. Many of these properties depend sensitively on SWCNT structure, which is characterized by the chiral index (n, m) that denotes the length and orientation of the circumferential vector in the hexagonal carbon lattice. Electronic properties are particularly strongly affected, with subtle structural changes switching tubes from metallic to semiconducting with various bandgaps. Monodisperse ‘single-chirality’ (that is, with a single (n, m) index) SWCNTs are thus needed to fully exploit their technological potential^{1,2}. Controlled synthesis through catalyst engineering^{3–6}, end-cap engineering⁷ or cloning strategies^{8,9}, and also tube sorting based on chromatography^{10,11}, density-gradient centrifugation, electrophoresis and other techniques¹², have delivered SWCNT samples with narrow distributions of tube diameter and a large fraction of a pre-determined tube type. But an effective pathway to truly monodisperse SWCNTs remains elusive. The use of template molecules to unambiguously dictate the diameter and chirality of the resulting nanotube^{8,13–16} holds great promise in this regard, but has hitherto had only limited practical success^{7,17,18}. Here we show that this bottom-up strategy can produce targeted nanotubes: we convert molecular precursors into ultrashort singly capped (6,6) ‘armchair’ nanotube seeds using surface-catalysed cyclodehydrogenation on a platinum (111) surface, and then elongate these during a subsequent growth phase to produce single-chirality and essentially defect-free SWCNTs with lengths up to a few hundred nanometres. We expect that our on-surface synthesis approach will provide a route to nanotube-based materials with highly optimized properties for applications such as light detectors, photovoltaics, field-effect transistors and sensors².

Recent work has produced non-planar carbon-based nanostructures such as fullerenes, carbon pyramids, and buckybowl from their corresponding quasi-planar polycyclic aromatic hydrocarbon precursors through surface-catalysed cyclodehydrogenation (CDH)^{19–21}. We have extended the methodology to the synthesis of ultrashort singly capped SWCNTs, that is, a SWCNT end cap with a short tube segment attached. Such molecules represent ideal seeds for subsequent epitaxial elongation into isomerically pure SWCNTs. Formally, this approach mimics the conventional synthesis of SWCNTs by the root-growth mechanism, in which nanotube growth starts by nucleation of an end-cap fragment on a metal nanoparticle²². The key point is to avoid uncontrolled, spontaneous nucleation of end caps by providing atomically precise ultrashort nanotube seeds which unambiguously dictate the chiral index of SWCNTs forming on epitaxial elongation.

Precursor **P1** ($C_{96}H_{54}$; Fig. 1) was designed and synthesized by multi-step organic synthesis to tackle this challenge (for details, see Methods). Upon intramolecular CDH it affords seed **S1**, an ultra-short singly capped (6,6) SWCNT bearing a carbon nanotube segment. The selective growth of (6,6) SWCNTs is illustrated in Fig. 1 and combines two steps: (1)

formation of seed **S1**, and (2) subsequent epitaxial elongation. The first step is realized by depositing precursor **P1** on a Pt(111) surface followed by annealing to 770 K under ultrahigh vacuum conditions to induce the surface-catalysed CDH reaction (Fig. 2a, b). The second step, epitaxial elongation, is achieved by the incorporation of carbon atoms originating from the surface-catalysed decomposition of a carbon feedstock gas (Fig. 3a–c).

Figure 2c shows a scanning tunnelling microscopy (STM) image acquired after depositing **P1** on Pt(111). No step decoration or island formation is observed. Interactions with surrounding molecules and step edges are thus largely suppressed and ensure subsequent unperturbed CDH of the precursor^{19,20}. For the majority of the as-deposited precursors, STM reveals a three-fold symmetric conformation. However, the intrinsic

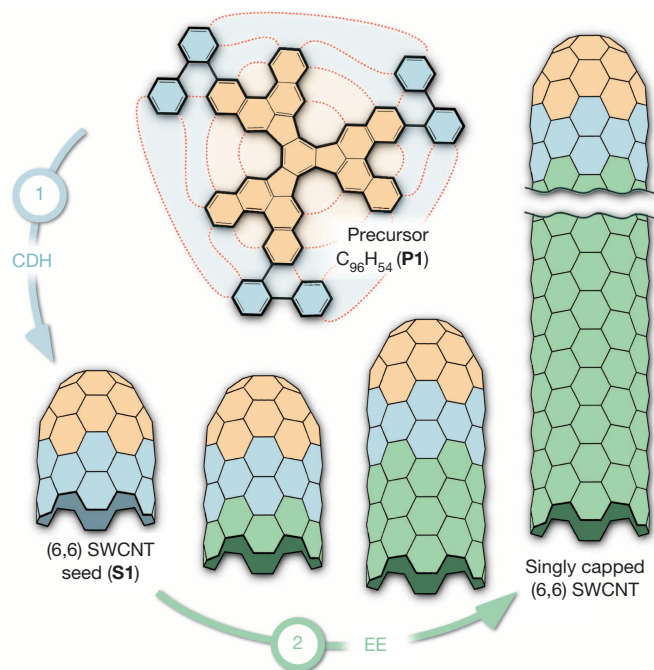


Figure 1 | Two-step bottom-up synthesis of SWCNTs. (1) Formation of singly capped ultrashort (6,6) SWCNT seed **S1** via cyclodehydrogenation (CDH) of the suitably designed polycyclic hydrocarbon precursor $C_{96}H_{54}$ (**P1**). (2) Nanotube growth via epitaxial elongation (EE). Parts of the precursor **P1** involved in the formation of the SWCNT end cap and the ultrashort CNT segment of the seed **S1** are highlighted in orange and blue, respectively. Red dashed lines indicate the new C–C bonds formed upon CDH. Epitaxial elongation occurs via the successive addition of carbon species, as indicated in green.

¹nanotech@surfaces Laboratory, Empa, Swiss Federal Laboratories for Materials Science and Technology, 8600 Dübendorf, Switzerland. ²Laboratory for Reliability Science and Technology, Empa, Swiss Federal Laboratories for Materials Science and Technology, 8600 Dübendorf, Switzerland. ³Max Planck Institute for Solid State Research, Heisenbergstrasse 1, 70569 Stuttgart, Germany. ⁴Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland. †Present addresses: Nanotechnology on Surfaces Laboratory, Instituto de Ciencia de Materiales de Sevilla (CSIC-US), Avenida Américo Vespucio 49, E-41092 Sevilla, Spain (J.R.S.-V.); BASF SE, GVM/I-L 544, 67056 Ludwigshafen, Germany (A.M.); University Erlangen-Nuremberg, Institut für Organische Chemie II, Henkestrasse 42, 91054 Erlangen, Germany (K.A.).

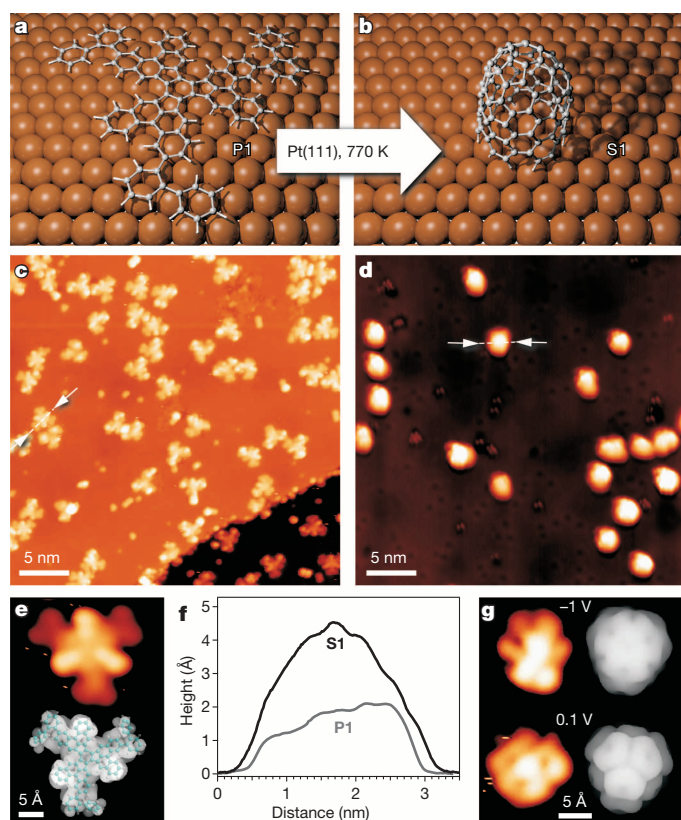


Figure 2 | Formation of (6,6) SWCNT seeds S1. **a, b,** Illustration of the thermally induced surface catalysed CDH to form the (6,6) SWCNT seed S1 from the adsorbed precursor P1. **c, d,** STM images of precursor molecules as deposited on Pt(111) (**c**) and after annealing to 770 K (**d**). **e,** Close-up STM image of a precursor (top) and the corresponding simulation based on the extended LUMO (greyscale) with a structural model of the molecule superimposed (bottom). **f,** Line profiles (positions indicated in **c, d**) over an as-deposited precursor P1 (grey line) and the seed species S1 obtained after annealing to 770 K (black line). **g,** Close-up STM images taken at -1 V and 0.1 V of (6,6) SWCNT seed S1 (left) and the corresponding simulations of HOMO and LUMO, respectively (greyscale, right).

axial chirality of the benzo[*c*]phenanthrene moieties ([4]helicene) (see Extended Data Fig. 1) and the configurational flexibility of the peripheral biphenyl moieties produce a large variety of possible geometries (Fig. 2c, Extended Data Fig. 1). The presence of axial chirality in the [4]helicene moieties results in four possible stereoisomers and thus eight possible adsorption geometries (Extended Data Fig. 1). An example of a molecule with its three outer biphenyl groups located closest to the surface is shown in Fig. 2e, together with the corresponding STM simulation based on the extended lowest unoccupied molecular orbital (LUMO; see Methods). The excellent agreement between STM image and simulation (Extended Data Fig. 1) indicates that the different topographic features observed for the adsorbed precursors can be attributed to the different adsorption geometries. Importantly, the stereoisomerism does not affect the CDH process, since all chiral centres will disappear during intramolecular cyclization.

Although P1 is designed to yield seed S1, the conformational flexibility of the peripheral biphenyl groups leads partially to undesired adsorption geometries. In contrast to the stereoisomers discussed above, these molecules will follow a different CDH pathway, ending in the formation of undesired buckybolls (Extended Data Fig. 2). A statistical analysis of more than 100 precursor monomers observed by STM revealed that more than 50% adopt the desired configurations (Extended Data Fig. 1). Most importantly, the condensation products of precursor molecules exhibiting 'wrong' conformations cannot act as seeds for the subsequent CNT growth process via epitaxial elongation, and thus will not affect the selectivity of SWCNT formation.

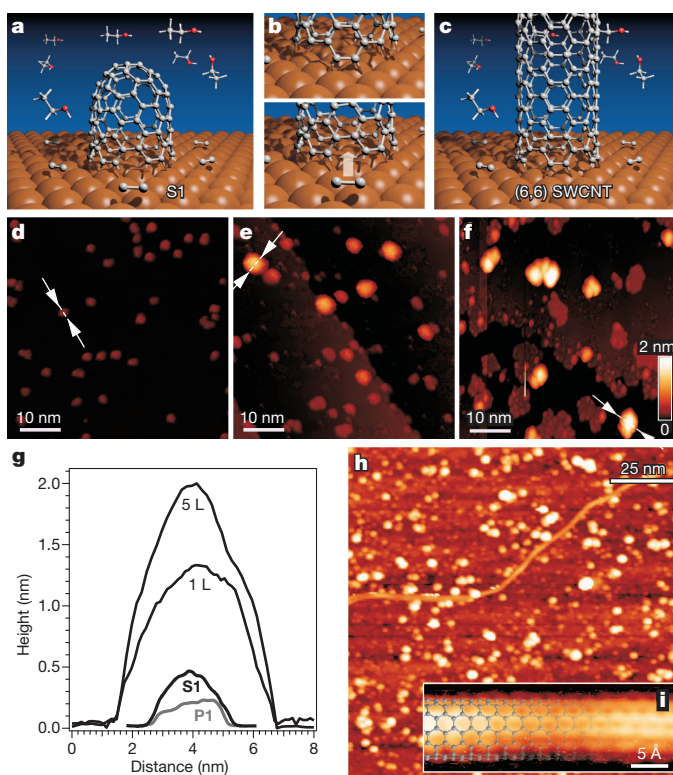


Figure 3 | Epitaxial elongation of singly capped SWCNT with (6,6) chiral index defined by the seed S1. **a–c,** Schematic illustration of the epitaxial elongation of (6,6) SWCNT seeds S1 via surface-catalysed C_2 incorporation at the nanotube/metal interface. The use of ethanol as a carbon feedstock gas is illustrated, which decomposes on the hot, reactive Pt surface into C_2 species and binds 'epitaxially' to the bay region of the previously formed singly capped SWCNT, resulting in an elongation of the tube along its axis. **d–f,** STM images of the as-prepared (6,6) SWCNT seeds (**d**), and after exposure to low doses of ethylene of 1 L (**e**) and 5 L (**f**), respectively, at a temperature of 770 K. For direct comparison, the same height colour scale is used for **d–f**. **g,** Line profiles taken across the features indicated by arrow pairs in the corresponding STM images **d–f**. **h, i,** STM images of a sample exposed to a pressure of 1×10^{-7} mbar of ethanol for 1 h (270 L) at a temperature of 770 K. A long SWCNT is observed to lie on the rough surface (**h**). A close-up STM image (**i**) identifies it as a (6,6) SWCNT.

Surface-catalysed CDH of precursors (P1) into seeds (S1) is induced by annealing at 770 K for 10 min. STM images (Fig. 2d) show that the originally quasi-planar three-fold symmetric molecules transform into dome-shaped species with a prominent increase in apparent height from 2 to 4.5 Å (Fig. 2f). Additional proof of successful dehydrogenation of P1 into S1 derives from the good agreement of high-resolution STM images and simulations of the frontier molecular orbitals of S1 (Fig. 2g). Both results demonstrate the successful formation of the targeted singly capped ultrashort (6,6) SWCNT S1.

In the second step, the surface-anchored seeds S1 are grown into (6,6) SWCNTs via epitaxial elongation by exposing them to a carbon feedstock gas such as ethylene or ethanol at temperatures between 670 and 770 K (see schematic illustration in Fig. 3a–c). The pre-synthesized seeds S1 are extended by catalytic epitaxial elongation, which consists of a consecutive incorporation of carbon atoms (schematically shown as C_2) originating from the decomposition of ethanol or ethylene. The open part of S1 is already in the required contact with the Pt(111) surface, which catalyses further epitaxial elongation. In order to unambiguously demonstrate the activity of S1 in an epitaxial elongation process, the results of low exposures to carbon feedstock gas were followed *in situ* by STM. Low doses of ethylene at 770 K produce a substantial increase in apparent height from 4 to 20 Å (Fig. 3d–g). After exposure to 1 Langmuir (L) of ethylene, about 18% of the initially deposited precursors P1 have

grown into SWCNTs, a density that remains constant for higher exposures of 5 L. Exposure to yet higher doses of ethylene or ethanol produces strong changes in the surface topography and makes STM imaging increasingly difficult (Fig. 3h). The topography becomes rough and the Pt surface is no longer discernible. Careful examination of the STM images reveals the presence of one-dimensional structures lying across the surface (Fig. 3h), with an abundance of about 5 per μm^2 and observed lengths exceeding 200 nm. High resolution images of the elongated structures (Fig. 3i) reveal an internal structure of higher contrast lines along the direction of the tube axis. A superimposed structural model clearly shows that these lines are indeed the carbon positions of the graphene structure in a (6,6) SWCNT. In all cases where atomic resolution of the tubes could be achieved, the structure proved to be consistent with a (6,6) SWCNT. The outstanding agreement of both the orientation and the periodicity of the graphene lattice with those expected demonstrate that the one-dimensional structures are the targeted (6,6) SWCNTs, bent horizontally across the sample surface.

To further corroborate the density and length of the horizontally aligned SWCNTs, the above sample was imaged with a scanning helium ion microscope (SHIM). Tubes longer than 300 nm and with diameters below 2 nm (the resolution limit) could be observed (Extended Data Fig. 3), with a density of 3–4 tubes per μm^2 that is similar to the density estimated from STM images (5 per μm^2). We note that both STM and SHIM can only image horizontally aligned SWCNTs, but do not give access to vertically aligned SWCNTs.

To shed light on the orientation of the SWCNTs and to characterize the selectivity of the growth process, Raman characterization was performed. Measurements with the illuminating laser beam at normal incidence (perpendicular to the surface plane) yielded extremely low intensities; the G band, the most intense feature at around $1,590\text{ cm}^{-1}$, is very weak (grey curve in Fig. 4a). However, when samples were measured under an illumination angle of 30° , all bands increased significantly in intensity (black curve in Fig. 4a), as expected for a dominant fraction of SWCNTs oriented perpendicular to the surface. The polarizability of these vertically oriented tubes with the illuminating laser beam at normal incidence—and thus the electromagnetic field vector perpendicular to the tube axis—is drastically reduced, resulting in weak Raman intensities²³. A predominantly vertical alignment of the SWCNTs is consistent with the observation that in SHIM images some CNTs appear to shake under the ion beam (Extended Data Fig. 3), and with the rough surface seen in STM. It also implies that the overall SWCNT density is expected to be significantly higher than that estimated from the STM and SHIM data.

More importantly, the Raman spectra demonstrate the high selectivity of our growth process, which produces (6,6) SWCNTs exclusively. The spectrum shown in Fig. 4b presents clearly defined bands at the positions expected for (6,6) SWCNTs. The band at 295 cm^{-1} is associated with the radial breathing mode (RBM) whose frequency depends strongly on the nanotube diameter²³. Empirical equations predict an RBM frequency of $280\text{--}295\text{ cm}^{-1}$ for (6,6) SWCNTs²³. Experimentally, the RBM for (6,6) SWCNTs deposited on SiO_2 has been reported at around 289 cm^{-1} (ref. 24), but it is well known that the substrate plays a crucial role in the RBM position, explaining the small deviation observed²⁵. Most importantly, our Raman spectra do not show any further bands within the RBM range ($200\text{--}400\text{ cm}^{-1}$), which underlines the extremely high selectivity of the process. Another notable characteristic of the RBM is the exceptionally small width observed (the full-width at half-maximum, FWHM, is 3.5 cm^{-1} , limited by the resolution of the instrument), which is as small as that from isolated individual small diameter SWCNTs (an FWHM of 3 cm^{-1})²⁶. The Raman spectra reported here average over a large area (beam diameter $1.5\text{--}10\text{ }\mu\text{m}$ (inset in Fig. 4b)), and thus reflect the properties of a large number of SWCNTs. The narrow RBM peak therefore demonstrates a very high degree of monodispersity.

The G band appears as a double peak at $1,518$ and $1,591\text{ cm}^{-1}$. The significant curvature in small diameter SWCNTs causes a shift to lower frequencies of the optical vibrations associated with transverse (perpendicular to the tube axis) atomic displacements (the G^- band)²⁷. Although,

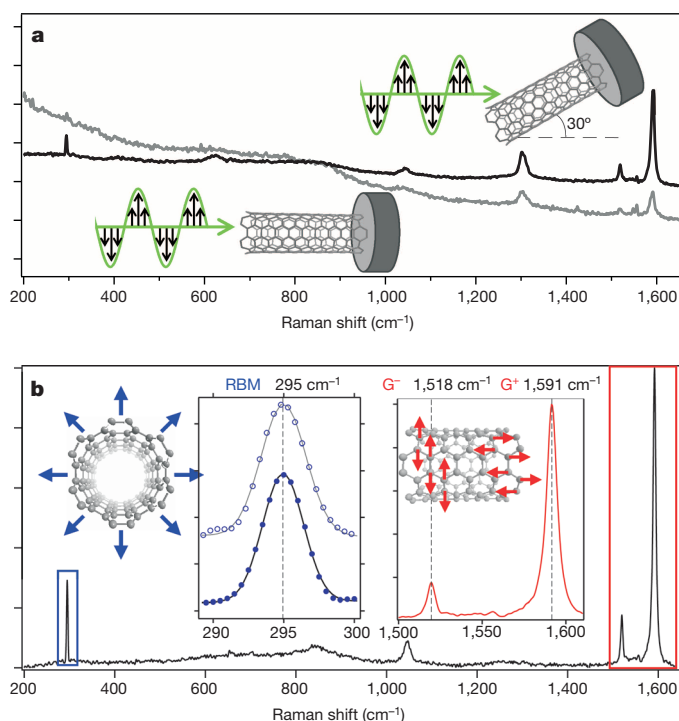


Figure 4 | SWCNT orientation determination and single chirality assessment by Raman spectroscopy. **a**, Raman spectra of epitaxially elongated (6,6) SWCNTs obtained by exposing seeds S1 to a pressure of 1×10^{-7} mbar of ethanol for 30 min (140 L) at a temperature of 670 K. The spectra were acquired with the sample surface perpendicular (grey curve) and at an angle of 30° (black curve) to the laser beam. **b**, Raman spectrum of longer SWCNTs (1 h at 1×10^{-7} mbar of ethanol at 670 K; 270 L) for a short laser illumination time (30 s), revealing defect-free SWCNTs as judged by the absence of a D peak. The insets show further details on the very narrow RBM at 295 cm^{-1} as measured with a beam diameter of 1.5 (filled circles) and $10\text{ }\mu\text{m}$ (open circles), respectively, and the splitting of the G band into a G^+ component at $1,591$ and a G^- component at $1,518\text{ cm}^{-1}$, which is characteristic of (6,6) SWCNTs.

to our knowledge, this splitting has not been observed experimentally for a (6,6) SWCNT, the G^+ and G^- band splitting ($\Delta\omega_G$) has been predicted to be 83 cm^{-1} (ref. 28), in good agreement with the splitting that we observe ($\Delta\omega_G = 73\text{ cm}^{-1}$). The additional peaks in the range from 400 to $1,200\text{ cm}^{-1}$ have previously been observed and used as evidence for the presence of armchair SWCNTs, since no peak in this range is present for semiconducting tubes²⁹. Finally, the absence of any D band in the Raman spectra (Fig. 4b) further underlines the extreme cleanliness of our process that yields not only predefined single-chirality but also essentially defect-free SWCNTs.

These findings clearly illustrate that the use of a planar metal surface instead of metal nanoparticles effectively suppresses spontaneous cap formation and that this, in combination with the use of a suitable precursor species to produce a desired cap, enables highly selective SWCNT fabrication. The entire *in situ* process and the low temperatures involved in both the CDH step (770 K) and the subsequent epitaxial elongation (670 K) are fully compatible with complementary metal oxide semiconductor (CMOS) technology, and our methods might therefore solve two pivotal challenges in the realization of CNT-based integrated circuits for digital electronics: to provide SWCNTs with identical electronic properties, and to integrate these SWCNTs in device architectures (which may be achieved by site-specific deposition of the molecular precursor and/or the catalyst film). However, further progress that aims to be technologically relevant requires a process with higher growth yields that ideally approach unity. While we see considerable scope for optimization of the present process (in terms of catalyst, temperatures, pressures, and so on), an alternative is to replace the substrate-catalysed epitaxial elongation step by a more efficient process, such as hot filament chemical

vapour deposition³⁰. This would also allow more flexibility in independently optimizing seed formation and tube elongation while simultaneously suppressing spontaneous SWCNT growth at both stages. Another logical extension of the process is to other SWCNTs where subtleties related to the chiral index of the seed may come into play.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 April; accepted 17 June 2014.

- Jorio, A., Dresselhaus, G. & Dresselhaus, M. S. *Carbon Nanotubes: Advanced Topics in the Synthesis, Structure, Properties, and Applications* (Springer, 2008).
- Jariwala, D., Sangwan, V. K., Lauhon, L. J., Marks, T. J. & Hersam, M. C. Carbon nanomaterials for electronics, optoelectronics, photovoltaics, and sensing. *Chem. Soc. Rev.* **42**, 2824–2860 (2013).
- Wang, H. *et al.* Selective synthesis of (9,8) single walled carbon nanotubes on cobalt incorporated TUD-1 catalysts. *J. Am. Chem. Soc.* **132**, 16747–16749 (2010).
- He, M. *et al.* Selective growth of SWNTs on partially reduced monometallic cobalt catalyst. *Chem. Commun.* **47**, 1219–1221 (2011).
- Chiang, W.-H. & Mohan Sankaran, R. Linking catalyst composition to chirality distributions of as-grown single-walled carbon nanotubes by tuning $\text{Ni}_x\text{Fe}_{1-x}$ nanoparticles. *Nature Mater.* **8**, 882–886 (2009).
- Hong, G., Chen, Y., Li, P. & Zhang, J. Controlling the growth of single-walled carbon nanotubes on surfaces using metal and non-metal catalysts. *Carbon* **50**, 2067–2082 (2012).
- Yu, X. *et al.* Cap formation engineering: from opened C_{60} to single-walled carbon nanotubes. *Nano Lett.* **10**, 3343–3349 (2010).
- Smalley, R. E. *et al.* Single wall carbon nanotube amplification: en route to a type-specific growth mechanism. *J. Am. Chem. Soc.* **128**, 15824–15829 (2006).
- Yao, Y., Feng, C., Zhang, J. & Liu, Z. ‘Cloning’ of single-walled carbon nanotubes via open-end growth mechanism. *Nano Lett.* **9**, 1673–1677 (2009).
- Tu, X., Manohar, S., Jagota, A. & Zheng, M. DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes. *Nature* **460**, 250–253 (2009).
- Liu, H., Nishide, D., Tanaka, T. & Kataura, H. Large-scale single-chirality separation of single-wall carbon nanotubes by simple gel chromatography. *Nature Commun.* **2**, 309 (2011).
- Hersam, M. C. Progress towards monodisperse single-walled carbon nanotubes. *Nature Nanotechnol.* **3**, 387–394 (2008).
- Mueller, A., Amsharov, K. Y. & Jansen, M. Synthesis of end-cap precursor molecules for (6, 6) armchair and (9, 0) zig-zag single-walled carbon nanotubes. *Tetrahedr. Lett.* **51**, 3221–3225 (2010).
- Bunz, U. H. F., Menning, S. & Martín, N. para-Connected cyclophenylenes and hemispherical polyarenes: building blocks for single-walled carbon nanotubes? *Angew. Chem. Int. Edn Engl.* **51**, 7094–7101 (2012).
- Mueller, A. & Amsharov, K. Y. Synthesis of precursors for large-diameter hemispherical buckybowls and precursors for short carbon nanotubes. *Eur. J. Org. Chem.* **2012**, 6155–6164 (2012).
- Omachi, H., Segawa, Y. & Itami, K. Synthesis of cycloparaphenylenes and related carbon nanorings: a step toward the controlled synthesis of carbon nanotubes. *Acc. Chem. Res.* **45**, 1378–1389 (2012).
- Mueller, A., Amsharov, K. Y. & Jansen, M. End-cap precursor molecules for the controlled growth of single-walled carbon nanotubes. *Fullerenes Nanotubes Carbon Nanostruct.* **20**, 401–404 (2012).
- Omachi, H., Nakayama, T., Takahashi, E., Segawa, Y. & Itami, K. Initiation of carbon nanotube growth by well-defined carbon nanorings. *Nature Chem.* **5**, 572–576 (2013).
- Otero, G. *et al.* Fullerenes from aromatic precursors by surface-catalysed cyclodehydrogenation. *Nature* **454**, 865–868 (2008).
- Amsharov, K. *et al.* Towards the isomer-specific synthesis of higher fullerenes and buckybowls by the surface-catalyzed cyclodehydrogenation of aromatic precursors. *Angew. Chem. Int. Edn Engl.* **49**, 9392–9396 (2010).
- Rim, K. T. *et al.* Forming aromatic hemispheres on transition-metal surfaces. *Angew. Chem. Int. Edn Engl.* **46**, 7891–7895 (2007).
- Gavillet, J. *et al.* Root-growth mechanism for single-wall carbon nanotubes. *Phys. Rev. Lett.* **87**, 275504 (2001).
- Dresselhaus, M. S., Dresselhaus, G., Saito, R. & Jorio, A. Raman spectroscopy of carbon nanotubes. *Phys. Rep.* **409**, 47–99 (2005).
- Jorio, A. *et al.* Quantifying carbon-nanotube species with resonance Raman scattering. *Phys. Rev. B* **72**, 075207 (2005).
- Soares, J. S. & Jorio, A. Study of carbon nanotube-substrate interaction. *J. Nanotechnol.* **2012**, 1–10 (2012).
- Jorio, A. *et al.* Linewidth of the Raman features of individual single-wall carbon nanotubes. *Phys. Rev. B* **66**, 115411 (2002).
- Telg, H. *et al.* Chiral index dependence of the G^+ and G^- Raman modes in semiconducting carbon nanotubes. *ACS Nano* **6**, 904–911 (2012).
- Piscanec, S., Lazzeri, M., Robertson, J., Ferrari, A. & Mauri, F. Optical phonons in carbon nanotubes: Kohn anomalies, Peierls distortions, and dynamic effects. *Phys. Rev. B* **75**, 035427 (2007).
- Rao, A. M. *et al.* Diameter-selective Raman scattering from vibrational modes in carbon nanotubes. *Science* **275**, 187–191 (1997).
- Xu, Y.-Q. *et al.* Vertical array growth of small diameter single-walled carbon nanotubes. *J. Am. Chem. Soc.* **128**, 6560–6561 (2006).

Acknowledgements This research was supported in part by the Swiss National Science Foundation and by the State Secretariat for Education, Research and Innovation via the COST Action MP0901 ‘NanoTP’. K.A. acknowledges financial support from Deutsche Forschungsgemeinschaft.

Author Contributions K.A., M.J. and R.F. initiated and conceived this work. K.A. designed the precursor molecules and the corresponding synthetic routes, K.A. and A.M. synthesised precursor molecules and performed HPLC, NMR and MS analyses. J.R.S.-V. carried out on-surface synthesis work. J.R.S.-V. and T.D. performed STM and Raman measurements, I.S. did the He ion microscopy analysis. O.G. performed the calculations. All authors participated in analysis and interpretation of the results. J.R.S.-V. drafted the manuscript, with contributions from P.R. and O.G. R.F. and K.A. edited the manuscript and coordinated the efforts of the research teams.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.F. (roman.fasel@empa.ch) or K.A. (konstantin.amsharov@fau.de).

A global ocean inventory of anthropogenic mercury based on water column measurements

Carl H. Lamborg¹, Chad R. Hammerschmidt², Katlin L. Bowman², Gretchen J. Swarr¹, Kathleen M. Munson¹, Daniel C. Ohnemus¹, Phoebe J. Lam¹, Lars-Eric Heimbürger³, Michä J. A. Rijkenberg⁴ & Mak A. Saito¹

Mercury is a toxic, bioaccumulating trace metal whose emissions to the environment have increased significantly as a result of anthropogenic activities such as mining and fossil fuel combustion^{1,2}. Several recent models have estimated that these emissions have increased the oceanic mercury inventory by 36–1,313 million moles since the 1500s^{2–9}. Such predictions have remained largely untested owing to a lack of appropriate historical data and natural archives. Here we report oceanographic measurements of total dissolved mercury and related parameters from several recent expeditions to the Atlantic, Pacific, Southern and Arctic oceans. We find that deep North Atlantic waters and most intermediate waters are anomalously enriched in mercury relative to the deep waters of the South Atlantic, Southern and Pacific oceans, probably as a result of the incorporation of anthropogenic mercury. We estimate the total amount of anthropogenic mercury present in the global ocean to be 290 ± 80 million moles, with almost two-thirds residing in water shallower than a thousand metres. Our findings suggest that anthropogenic perturbations to the global mercury cycle have led to an approximately 150 per cent increase in the amount of mercury in thermocline waters and have tripled the mercury content of surface waters compared to pre-anthropogenic conditions. This information may aid our understanding of the processes and the depths at which inorganic mercury species are converted into toxic methyl mercury and subsequently bioaccumulated in marine food webs.

Mercury (Hg) is emitted to the atmosphere by natural and human sources primarily as Hg^0 , which is unusually volatile for a metal¹. The elemental form is removed from the atmosphere after oxidation to Hg^{2+} , which is then deposited to land and ocean. Within the ocean, Hg^{2+} is readily reduced to Hg^0 , resulting in surface waters being supersaturated in the elemental form with respect to the atmosphere. With an atmospheric lifetime of between a few months and a year as well as the evasion of Hg^0 from the ocean to the atmosphere, Hg from any source can

be widely dispersed across the globe⁵. Hg in the ocean is also subject to bioaccumulation and scavenging by organic-rich particles. Such particles eventually sink out of the surface ocean and are respired at deeper depths, transporting carbon, nutrients and metals like Hg in the process. In this way, Hg is very much like carbon dioxide (CO_2) in that it is a biologically active gas that exhibits wide dispersal in the atmosphere, vigorous air–sea exchange and vertical transport in the ocean as a result of the particulate “biological carbon pump”¹⁰. Like the other group-12 elements (Zn and Cd)¹¹ we might expect Hg distributions in the ocean to mimic macronutrients like phosphate (PO_4^{3-}) (low in the surface, increasing through the thermocline, higher in the deep Pacific than in the deep Atlantic). As can be seen in some representative vertical profiles of Hg concentrations (Fig. 1), this general trend is indeed observed.

However, oceanic Hg distributions are a combination of pre-anthropogenic, nutrient-like and transient signals resulting from human activities over the past several centuries. Figure 2 shows the concentrations of Hg and dissolved phosphate released during organic matter remineralization (P_{remin} is the apparent oxygen utilization divided by 170; ref. 12) measured in a variety of water masses from GEOTRACES cruises to the North and South Atlantic Ocean, the Pacific sector of the Southern Ocean, a GEOTRACES Inter calibration cruise to the subtropical northeast Pacific Ocean and non-GEOTRACES cruises to the tropical Pacific Ocean (the ‘Metalloenzyme’ cruise), the North Pacific Ocean (CLIVAR Repeat P16), and the central Arctic Ocean (2011 Polarstern cruise ARK-XXVI/3–TransArc) (refs 13–16 and K.M.M., C.H.L., G.J.S. & M.A.S., manuscript in preparation, but unpublished data available at <http://www.bco-dmo.org> or on request). In the water masses other than Northern Hemisphere North Atlantic Deep Water and recently subducted Antarctic Bottom Water (henceforth referred to as ‘unaffected’ deep waters), a striking correlation between Hg and P_{remin} is seen (the reduced major axis regression line in Fig. 2). This correlation offers several important insights: (1) these water masses possess little anthropogenic

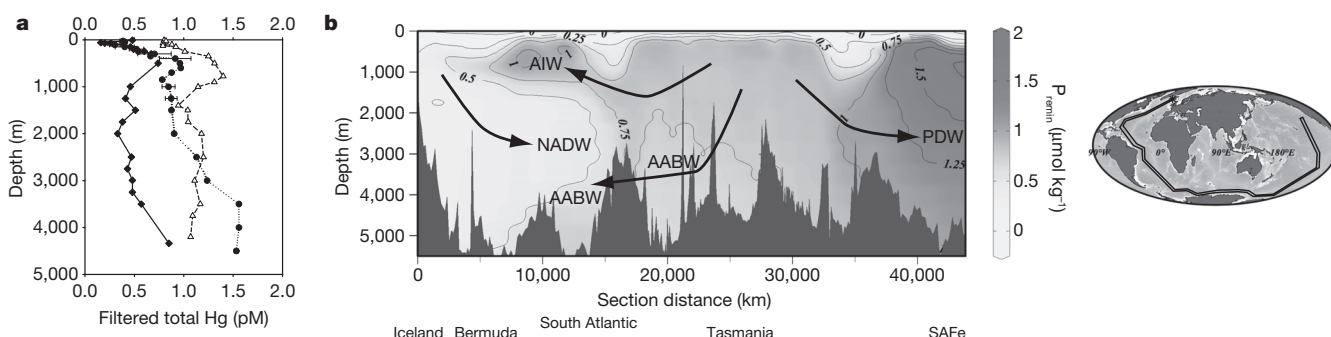


Figure 1 | Hg and P_{remin} distributions in the ocean. **a**, Representative profiles of total dissolved Hg from the North Atlantic near Bermuda (triangles), from the South Atlantic (diamonds) and the northeast Pacific Ocean (circles; the station labelled ‘SAFe’). **b**, Global vertical distribution of P_{remin} as interpolated

from WOCE data (<http://www.ewoce.org>). Transect (black lines) is shown to the right. Figure generated using Ocean Data View (<http://odv.awi.de/>). AIW, Atlantic Intermediate water; NADW, North Atlantic Deep Water; AABW, Antarctic Bottom Water; PDW, Pacific Deep Water.

¹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. ²Department of Earth and Environmental Sciences, Wright State University, Dayton, Ohio 45435, USA. ³Observatoire Midi-Pyrénées, Laboratoire Géosciences Environnement Toulouse, CNRS/IRD/Université Paul-Sabatier, 14, avenue Édouard Belin, 31400 Toulouse, France. ⁴Department of Biological Oceanography, Royal Netherlands Institute for Sea Research, Den Burg, 1790 AB, The Netherlands.

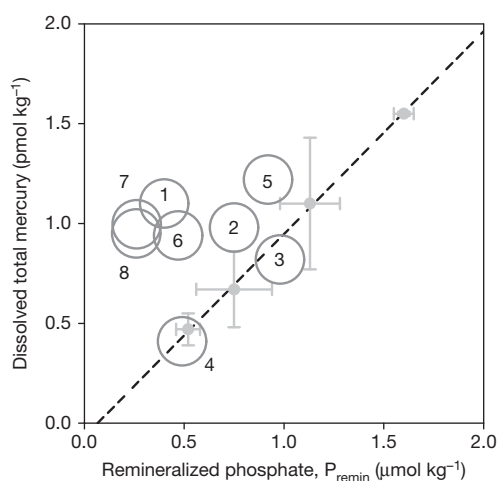


Figure 2 | The concentration of Hg and P_{remin} in various water masses. The grey symbols are the data from deep waters (>1,000 m depth) not suspected of containing anthropogenic Hg. The remaining symbols include North Atlantic Deep Water in the North Atlantic Ocean (1), Antarctic Bottom Water sampled between Tasmania and Antarctica (2), thermocline waters from the North Atlantic Ocean (6), South Atlantic Ocean (4), the Southern Ocean between Tasmania and Antarctica (8), the Arctic Ocean (7), the northeast Pacific Ocean (5) and the tropical Pacific Ocean (3).

Hg delivered by the biological pump, otherwise a good correlation and a y -intercept that is almost zero ($-0.07 \pm 0.03 \text{ pmol kg}^{-1} \text{ Hg}$) would not have been observed (see Supplementary Information); (2) the slope of the line is an expression of the Hg/P ratio in sinking organic matter formed in surface waters from before the anthropogenic impact ($1.02 \pm 0.03 \text{ } \mu\text{mol Hg per mole P}$); (3) the relationship between P_{remin} and Hg allows us to use it as a benchmark against which water masses that do contain anthropogenic Hg can be compared.

The impact of anthropogenic Hg emissions in the deep North Atlantic and various thermocline water masses is evident in Fig. 2, with data points that lie above the unaffected deep water regression line, thus showing evidence of anthropogenic Hg contributions; the vertical distance between the data and the line represents the amount of Hg in that water mass contributed from human sources. It is immediately apparent, however, that the degree of Hg perturbation for each water mass is not the same. This can be explored further by dividing the amount of anthropogenic Hg in each water mass by a tracer, preferably a pollutant that has a similar emissions history. This will allow the derived amount of anthropogenic mercury (Hg_{anth}) to be cross-checked against expectations as well as greatly simplify the conversion of our measurements to a scaled-up estimate of the total amount of pollutant Hg in the ocean. For this purpose, we have selected the amount of anthropogenic carbon dioxide ($\text{CO}_{2,\text{anth}}$) present in each water mass (Table 1). The $\text{CO}_{2,\text{anth}}$ estimates were derived using the ΔC^* method of Gruber and colleagues¹⁷ from a variety of data sets¹⁸, and then gridded over the whole ocean (the GLODAP database)¹⁹. The $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratios in most of the water masses are not statistically different from either each other or the Hg/ CO_2 ratio in primary anthropogenic atmospheric emissions ($9.6\text{--}12.4 \text{ Mmol Hg per year}$; $0.79 \pm 0.04 \text{ Pmol C per year}$; $\text{Hg/C} = 14 \pm 2 \text{ nmol mol}^{-1}$)^{20–22}. However, shallower water masses appear to have smaller mean $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratios than either North Atlantic Deep Water (NADW) or Antarctic Bottom Water, which have mean $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratios exceeding those in most known emissions sources²¹. The cause of this higher ratio is unclear, but it may be attributed to either high localized rates of atmospheric Hg deposition due to high rates of precipitation (Southern Ocean), enrichment caused by salt rejection during sea-ice formation¹⁴, proximity to historically strong regions of Hg emissions in North America and Europe (North Atlantic) or the prevalence of coal burning as a source of CO_2 early in the Industrial Revolution. For example, surface waters near Iceland (the site of NADW formation) and Antarctica (Antarctic Bottom Water) are enriched in Hg (about 2 pM)^{14,23} with respect to average surface waters (0.6 pM ; see

Table 1 | Summary of Hg, P_{remin} and $\text{CO}_{2,\text{anth}}$ data

Ocean basin; water mass	Hg (pmol kg^{-1})	P_{remin} ($\mu\text{mol kg}^{-1}$)	Hg_{anth} (pmol kg^{-1})	$\text{CO}_{2,\text{anth}}$ ($\mu\text{mol kg}^{-1}$)	(Hg/CO_2) _{anth} (nmol mol^{-1})	Selection criteria (stations; depth)
Unaffected deep waters						
South Atlantic; North Atlantic Deep Water	0.47 ± 0.08	0.52 ± 0.06	0	0	NA	All stations; 1,500–4,000 m
South Atlantic; Antarctic Bottom Water	0.67 ± 0.19	0.75 ± 0.19	0	0	NA	All stations; below 4,000 m
Tropical Pacific; Pacific Deep Water and Pacific Bottom Water	1.10 ± 0.33	1.13 ± 0.15	0	0	NA	All stations; below 1,000 m
Subtropical northeast Pacific; Pacific Deep Water and Pacific Bottom Water	1.55 ± 0.01	1.60 ± 0.05	0	0	NA	One station; below 1,000 m
Affected deep waters						
North Atlantic; North Atlantic Deep Water	1.1 ± 0.2	0.40 ± 0.05	0.72 ± 0.17	10.2 ± 9.7	58 ± 29	All stations; 1,500–4,000 m
Southern Ocean; Antarctic Bottom Water	0.98 ± 0.17	0.75 ± 0.10	0.29 ± 0.20	5 ± 4	76 ± 8	All stations south of 50° S , depths 100–1,000 m
Thermocline waters						
South Atlantic; thermocline	0.41 ± 0.14	0.49 ± 0.35	-0.02 ± 0.28	25 ± 13	NA	All stations; 100–1,000 m
Tropical Pacific; thermocline	0.82 ± 0.35	0.98 ± 0.48	-0.10 ± 0.60	16 ± 16	NA	All stations; 100–1,000 m
North Atlantic; thermocline	0.94 ± 0.27	0.47 ± 0.30	0.52 ± 0.21	42 ± 14	15 ± 8	All stations; 100–1,000 m
Northeast Pacific; thermocline	1.22 ± 0.39	0.92 ± 0.72	0.35 ± 0.59	25 ± 14	23 ± 18	All stations; 100–1,000 m
Arctic; thermocline	1.00 ± 0.11	0.26 ± 0.03	0.80 ± 0.12	~ 30	~ 27	Two stations; 200–1,000 m
Southern Ocean; thermocline	0.95 ± 0.057	0.47 ± 0.33	0.66 ± 0.57	22 ± 12	37 ± 24	All stations south of 50° S , depths 100–1,000 m

Hg, P_{remin} and $\text{CO}_{2,\text{anth}}$ values are water mass averages. NA, not applicable. The Hg_{anth} and (Hg/CO_2)_{anth} values shown are averages of sample-by-sample calculations.

below), which is consistent with greater mean Hg/CO_2 ratios in these deep waters. Some alteration in $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratios should also be expected from the differential behaviours in the ocean between these two biologically active gases (Hg_{anth} will be moved into the thermocline and mode waters by both the biological as well as the solubility pump¹⁰, whereas $\text{CO}_{2,\text{anth}}$ will not be pumped biologically because oceanic primary productivity is not C-limited).

We used the observed $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratios in each affected water mass to estimate the inventory of Hg_{anth} in the ocean as a whole by multiplying these ratios by the estimated amount of $\text{CO}_{2,\text{anth}}$ in the ocean ($9.8 \pm 1.6 \text{ Pmol C}$)¹⁹. Given the still small and evolving amount of oceanographic Hg data available, we chose to use one $\text{Hg}_{\text{anth}}/\text{CO}_{2,\text{anth}}$ ratio for intermediate waters ($100\text{--}1,000 \text{ m}$; $25 \pm 11 \text{ nmol mol}^{-1}$) and another for the deep North Atlantic ($66 \pm 14 \text{ nmol mol}^{-1}$); we used the GLODAP model estimate for the percentage of $\text{CO}_{2,\text{anth}}$ in each ocean layer: 15% in surface water, 71% in intermediate waters, and 16% in deep water. This calculation suggests that there are about $170 \pm 80 \text{ Mmol}$ of anthropogenic Hg between 100 m and 1,000 m depth and about $100 \pm 20 \text{ Mmol}$ deeper than 1,000 m.

It is not appropriate to use P_{remin} to identify the anthropogenic impact on Hg in waters shallower than 100 m because atmospheric deposition is the primary source of Hg to the surface ocean, not particle remineralization. Alternatively, we estimated Hg_{anth} in surface waters by comparing the slope of the regression in Fig. 2 with the Hg/P ratio in contemporary suspended particulate matter. The Hg/P ratio was derived from analysis of Hg and P in mixed-layer particulate matter collected by *in situ* pumping performed during both the North Atlantic 'GEOTRACES' and tropical Pacific 'Metalloenzyme' cruises. This ratio is $3.4 \pm 1.3 \mu\text{mol Hg per mole P}$, indicating a factor of 3.4 ± 1.3 increase in the concentration of Hg in surface ocean particulate matter and presumably in solution as well since industrialization. This degree of secular change of Hg in surface waters is consistent with archives of atmospheric Hg deposition that indicate a 2–5-fold increase worldwide since industrialization²⁴. The data presented here suggest that the total amount of Hg in the top 100 m of the ocean is about 22 Mmol (an average concentration of 0.6 pM). Accordingly, Hg_{anth} in this layer is about $16 \pm 6 \text{ Mmol}$.

Our overall estimate of $290 \pm 80 \text{ Mmol}$ (rounded to two significant figures) of Hg_{anth} in the ocean is in reasonable agreement with a number of model-based predictions^{4,7,8,25}, but suggests that the highest and lowest model estimates are implausible. On the high end is the prediction of Streets and colleagues², who estimated an amount of Hg_{anth} in the ocean of 1,313 Mmol, which required a major contribution from artisanal and small-scale gold mining at present and in the past. It is important to test this particular inventory² because it has featured prominently in recent negotiations concerning international efforts to curb emissions of Hg to the environment²⁶. Our measurements and calculations here suggest that either the Streets² estimate for past Hg anthropogenic releases is too high, or that much of the Hg they predicted to be in the ocean resides elsewhere, such as in soils. Recent work by Jaegle, Zhang and colleagues²⁷ has provided support for this as well by using modelling fits to water column profiles that also suggest that loadings to the ocean are lower than those of Streets and colleagues². It should be noted that the estimate for total $\text{CO}_{2,\text{anth}}$ to which we have indexed¹⁸ is for the year 1994. Estimates for more recent times and with different methods suggest greater $\text{CO}_{2,\text{anth}}$ (for example, 12.9 Pmol; ref. 28) which would predict higher values of Hg_{anth} as well (380 Mmol). However, this higher estimate is still much less than that of ref. 2.

As noted, we found that about a third of anthropogenic Hg loadings to the ocean are in deep water, particularly NADW. One model with which our results agree quite well is that of Sunderland and Mason⁷, who used a multi-box model that explicitly included deep water formation in the North Atlantic. In their simulation, 129 Mmol of Hg_{anth} are in ocean water shallower than 1,500 m, with another 124 Mmol in deeper waters. Thus, the prevalence of anthropogenic Hg in deep waters of the North Atlantic indicate the importance, as captured by the Sunderland

and Mason model⁷, of deep water formation for sequestration of surface Hg on millennial timescales. This observation also leads to the conclusion, given that Hg emissions from anthropogenic sources are predicted to increase at a rate faster than in the previous few centuries²⁰, that future loadings may somewhat overwhelm the deep water formation sink. We should therefore expect that the rate of increase of Hg in surface waters in the next few decades will be greater than the rate of increase in emissions during the same time period.

The impact of anthropogenic loadings on the oceanic Hg reservoir can be estimated with knowledge of the total amount of Hg in the ocean. Taking the North and South Atlantic concentration profiles each to represent a quarter of the whole ocean and the Pacific profiles to represent the other half, we estimated that the ocean contains 1,390 Mmol of dissolved total Hg, with 22 Mmol in the 0–100 m surface ocean, 292 Mmol in the 100–1,000 m intermediate depths and 1,260 Mmol in waters deeper than 1,000 m (the average concentration in these three layers being 0.6 pM, 0.9 pM and 1.0 pM, respectively^{13–16} (G.J.S., C.H.L., M.J.A.R. & C.R.H., manuscript in preparation)). These amounts are smaller than most previous estimates; for example, Sunderland and Mason⁷ estimated 666 Mmol in water shallower than 1,500 m and 1,095 Mmol in deeper water. Thus, analysis of the new data presented here suggests that the relative impact of human Hg emissions on the ocean is greater than previously thought: waters shallower than 1,000 m appear to have contained 120 Mmol in the pre-industrial past, and exhibit a factor-of-2.6 increase, while the ocean as a whole has experienced a factor-of-1.1 increase.

As our analysis reveals, and as has been noted elsewhere²⁵, the impact of human Hg emissions is not uniform within the ocean. Therefore, the extent to which methyl mercury concentrations in fish have changed since industrialization, and might change in response to further perturbation (perhaps as much as a fivefold increase over pre-industrial levels by 2050)²¹ can be determined only following studies of the vertical patterns in Hg methylation dynamics as well as basin-scale controls on methylation of anthropogenic Hg.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 May 2013; accepted 3 June 2014.

1. Fitzgerald, W. F. & Lamborg, C. H. in *Treatise on Geochemistry* (eds Holland, H. D. & Turekian, K. K.) Vol. 9, Ch. 4, 1–47 (Pergamon, 2003).
2. Streets, D. G. *et al.* All-time releases of mercury to the atmosphere from human activities. *Environ. Sci. Technol.* **45**, 10485–10491 (2011).
3. Mason, R. P., Fitzgerald, W. F. & Morel, F. M. M. The biogeochemical cycling of elemental mercury—anthropogenic influences. *Geochim. Cosmochim. Acta* **58**, 3191–3198 (1994).
4. Lamborg, C. H., Fitzgerald, W. F., O'Donnell, J. & Torgersen, T. A non-steady-state compartmental model of global-scale mercury biogeochemistry with interhemispheric atmospheric gradients. *Geochim. Cosmochim. Acta* **66**, 1105–1118 (2002).
5. Selin, N. E. *et al.* Global 3-D land-ocean-atmosphere model for mercury: present-day versus preindustrial cycles and anthropogenic enrichment factors for deposition. *Glob. Biogeochem. Cycles* **22**, GB2011 (2008).
6. Soerensen, A. L. *et al.* An improved global model for air-sea exchange of mercury: high concentrations over the North Atlantic. *Environ. Sci. Technol.* **44**, 8574–8580 (2010).
7. Sunderland, E. M. & Mason, R. P. Human impacts on open ocean mercury concentrations. *Glob. Biogeochem. Cycles* **21**, GB4022 (2007).
8. Strode, S., Jaegle, L. & Emerson, S. Vertical transport of anthropogenic mercury in the ocean. *Glob. Biogeochem. Cycles* **24**, GB4014 (2010).
9. Amos, H. M., Jacob, D. J., Streets, D. G. & Sunderland, E. M. Legacy impacts of all-time anthropogenic emissions on the global mercury cycle. *Glob. Biogeochem. Cycles* **27**, 410–421 (2013).
10. Volk, T. & Hoffert, M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. & Broecker, W. S.) 99–110 (American Geophysical Union, 1985).
11. Morel, F. M. M., Milligan, A. J. & Saito, M. A. in *Treatise on Geochemistry* Vol. 6 *The Oceans and Marine Geochemistry* (ed. Elderfield, H.) 113–143 (Elsevier, 2003).
12. Anderson, L. A. & Sarmiento, J. L. Redfield ratios of remineralization determined by nutrient data-analysis. *Glob. Biogeochem. Cycles* **8**, 65–80 (1994).
13. Bowman, K. L., Hammerschmidt, C. R., Lamborg, C. H. & Swarr, G. J. Mercury in the North Atlantic Ocean: the U.S. GEOTRACES zonal and meridional sections. *Deep Sea Res. II* (in the press).

14. Cossa, D. *et al.* Mercury in the Southern Ocean. *Geochim. Cosmochim. Acta* **75**, 4037–4052 (2011).
15. Hammerschmidt, C. R. & Bowman, K. L. Vertical methylmercury distribution in the subtropical North Pacific. *Mar. Chem.* **132–133**, 77–82 (2012).
16. Sunderland, E. M., Krabbenhoft, D. P., Moreau, J. W., Strode, S. A. & Landing, W. M. Mercury sources, distribution, and bioavailability in the North Pacific Ocean: insights from data and models. *Glob. Biogeochem. Cycles* **23**, GB2010 (2009).
17. Gruber, N., Sarmiento, J. L. & Stocker, T. F. An improved method for detecting anthropogenic CO₂ in the oceans. *Glob. Biogeochem. Cycles* **10**, 809–837 (1996).
18. Sabine, C. L. *et al.* The oceanic sink for anthropogenic CO₂. *Science* **305**, 367–371 (2004).
19. Key, R. M. *et al.* A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, GB4031 (2004).
20. Streets, D. G., Zhang, Q. & Wu, Y. Projections of global mercury emissions in 2050. *Environ. Sci. Technol.* **43**, 2983–2988 (2009).
21. Pacyna, E. G. *et al.* Global emission of mercury to the atmosphere from anthropogenic sources in 2005 and projections to 2020. *Atmos. Environ.* **44**, 2487–2499 (2010).
22. Le Quéré, C. *et al.* The global carbon budget 1959–2011. *Earth Syst. Sci. Data Discuss.* **5**, 1107–1157 (2012).
23. Mason, R. P., Rolffus, K. R. & Fitzgerald, W. F. Mercury in the North Atlantic. *Mar. Chem.* **61**, 37–53 (1998).
24. Lamborg, C. H. *et al.* Modern and historic atmospheric mercury fluxes in both hemispheres: global and regional mercury cycling implications. *Glob. Biogeochem. Cycles* **16**, 1104 (2002).
25. Mason, R. P. *et al.* Mercury biogeochemical cycling in the ocean and policy implications. *Environ. Res.* **119**, 101–117 (2012).
26. Selin, N. E. Global change and mercury cycling: challenges for implementing a global mercury treaty. *Environ. Toxicol. Chem.* **33**, 1202–1210 (2014).
27. Jaegle, L., Zhang, Y., Thompson, L., Emerson, S. & Trossman, D. The Past 600 Years: Changing Hg Concentrations in a Global 3D Ocean Tracer Model. In *11th Int. Conf. on Mercury as a Global Pollutant* (Edinburgh, 2013); abstr. M4-1000, <http://www.mercury2013.com/>.
28. Khatiwala, S. *et al.* Global ocean storage of anthropogenic carbon. *Biogeosciences* **10**, 2169–2191 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the captains and crews of all cruises, as well as: P. Morton, J. Fitzsimmons, R. Shelley, A. Aguilar-Islas, R. Bundy, P. Morris, S. Owens, K. Wang, S. Rigaud and S. Pike for sample collection during the North Atlantic GEOTRACES cruise; L. Groot, D. Weiss, P. Laan, J. de Jong, R. Middag, L. Pena, A. Hartman, J. M. Godoy, L. Gerringa, M. Boyé and J. Dérot for sample collection during the South Atlantic GEOTRACES cruise; T. Goepfert, E. Bertrand and D. Moran for sampling during the Metalloenzyme cruise; and M. Rutgers van der Loeff and B. Galfond for providing samples from the 2011 Polarstern cruise ARK-XXVI/3–TransArc to the central Arctic Ocean. We are also grateful to D. Cossa and E. Sunderland for providing digital versions of their Southern Ocean and P16 data. We also thank H. Amos, L. Jaegle, B. Jonsson, R. Mason, E. Sunderland and Y. Zhang for discussions and D. Cossa for comments. This work was supported by NSF grant numbers OCE-0825108, OCE-0825157, OCE-0927274, OCE-0928191, OCE-1031271, OCE-1132480 and OCE-1132515. We thank co-Principal Investigators R. Mason and G. Gill. L.-E.H. thanks J. E. Sonke for funding Arctic Ocean observations via research grant ERC-2010-StG_20091028 to JES.

Author Contributions C.H.L., C.R.H., K.L.B., G.J.S., D.C.O., L.-E.H., M.J.A.R. and M.A.S. participated in the GEOTRACES, Metalloenzyme cruises. C.H.L., C.R.H., K.L.B., G.J.S. and L.-E.H. performed Hg analyses. D.C.O. and P.J.L. designed the particulate sampling experiments and performed P analyses. C.H.L., C.R.H., M.J.A.R., M.A.S. and L.-E.H. designed the Hg-related experiments. C.H.L., C.R.H., K.L.B., G.J.S., K.M.M. and L.-E.H. interpreted the data. All authors contributed to manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.H.L. (clamborg@whoi.edu).

METHODS

Water samples were collected using ultraclean techniques²⁹, including the use of a largely metal-free collection system and pressure filtration to 0.45 µm of water samples directly from the sampling GO-Flo bottles. Aliquots for total 'dissolved' Hg were collected in 250-ml, acid-washed, borosilicate glass bottles, oxidized with BrCl and analysed by cold vapour atomic fluorescence spectrometry following SnCl₂ reduction and gold-trap pre-concentration^{30–32}.

P_{remin} was calculated according to Anderson and Sarmiento¹² as the apparent oxygen utilization divided by 170 ± 10 , where the apparent oxygen utilization is calculated as $[O_2]_{\text{saturated}} - [O_2]_{\text{observed}}$, where $[O_2]_{\text{saturated}}$ is determined from depth, temperature and salinity³³.

Particulate Hg and P were determined from subsamples of quartz-fibre or polyethersulphone filters loaded with suspended matter (<51 µm) using McLane pumps. For Hg, the filter aliquots were digested with 2 M HNO₃, and the digest treated as dissolved samples³⁴. For P, polyethersulphone filter subsamples were digested in a 3:1 sulphuric acid:hydrogen peroxide solution to oxidize and dissolve the polyethersulphone filter, dried down, and then particles were digested in a 4 M HCl/HNO₃/HF mixture³⁵. The digest was analysed for multiple elements including P on a high-resolution inductively coupled plasma mass spectrometer and standardized using multi-element external standards (similar to ref. 36).

Water masses were defined primarily based on depth (as noted in Table 1), in accordance with those suggested by Talley and colleagues³⁷. This definition represents

an approximation for more refined definitions made on the basis of temperature, salinity and basin.

29. Cutter, G. A. & Bruland, K. W. Rapid and noncontaminating sampling system for trace elements in global ocean surveys. *Limnol. Oceanogr. Meth.* **10**, 425–436 (2012).
30. Gill, G. A. & Fitzgerald, W. F. Picomolar mercury measurements in sea water and other materials using stannous chloride reduction and two-stage gold amalgamation with gas phase detection. *Mar. Chem.* **20**, 227–243 (1987).
31. Fitzgerald, W. F. & Gill, G. A. Subnanogram determination of mercury by two-stage gold amalgamation applied to atmospheric analysis. *Anal. Chem.* **51**, 1714–1720 (1979).
32. Lamborg, C. H., Hammerschmidt, C. R., Gill, G. A., Mason, R. P. & Gichuki, S. An intercomparison of procedures for the determination of total mercury in seawater and recommendations regarding mercury speciation during GEOTRACES cruises. *Limnol. Oceanogr. Meth.* **10**, 90–100 (2012).
33. Weiss, R. F. Solubility of nitrogen, oxygen and argon in water and seawater. *Deep-Sea Res.* **17**, 721–735 (1970).
34. Hammerschmidt, C. R. & Fitzgerald, W. F. Bioaccumulation and trophic transfer of methylmercury in Long Island Sound. *Arch. Environ. Contam. Toxicol.* **51**, 416–424 (2006).
35. Eggemann, D. W. & Betzer, P. R. Decomposition and analysis of refractory oceanic suspended materials. *Anal. Chem.* **48**, 886–890 (1976).
36. Lamborg, C. H., *et al.* The flux of bio- and lithogenic material associated with sinking particles in the mesopelagic "twilight zone" of the northwest and North Central Pacific Ocean. *Deep-Sea Res. II* **55**, 1540–1563 (2008).
37. Talley, L. D., *et al.* North Pacific Intermediate Water in the Kuroshio Oyashio mixed water region. *J. Phys. Oceanogr.* **25**, 475–501 (1995).

Negative regulation of the NLRP3 inflammasome by A20 protects against arthritis

Lieselotte Vande Walle^{1,2}, Nina Van Opdenbosch^{1,2}, Peggy Jacques³, Amelie Fossoul^{1,2}, Eveline Verheugen³, Peter Vogel⁴, Rudi Beyaert^{5,6}, Dirk Elewaut³, Thirumala-Devi Kanneganti⁴, Geert van Loo^{5,6*} & Mohamed Lamkanfi^{1,2*}

Rheumatoid arthritis is a chronic autoinflammatory disease that affects 1–2% of the world's population and is characterized by widespread joint inflammation. Interleukin-1 is an important mediator of cartilage destruction in rheumatic diseases¹, but our understanding of the upstream mechanisms leading to production of interleukin-1 β in rheumatoid arthritis is limited by the absence of suitable mouse models of the disease in which inflammasomes contribute to pathology. Myeloid-cell-specific deletion of the rheumatoid arthritis susceptibility gene *A20/Tnfrsf3* in mice (*A20^{myel-KO}* mice) triggers a spontaneous erosive polyarthritis that resembles rheumatoid arthritis in patients². Rheumatoid arthritis in *A20^{myel-KO}* mice is not rescued by deletion of tumour necrosis factor receptor 1 (ref. 2). Here we show, however, that it crucially relies on the Nlrp3 inflammasome and interleukin-1 receptor signalling. Macrophages lacking A20 have increased basal and lipopolysaccharide-induced expression levels of the inflammasome adaptor Nlrp3 and proIL-1 β . As a result, A20-deficiency in macrophages significantly enhances Nlrp3 inflammasome-mediated caspase-1 activation, pyroptosis and interleukin-1 β secretion by soluble and crystalline Nlrp3 stimuli. In contrast, activation of the Nlrp4 and AIM2 inflammasomes is not altered. Importantly, increased Nlrp3 inflammasome activation contributes to the pathology of rheumatoid arthritis *in vivo*, because deletion of Nlrp3, caspase-1 and the interleukin-1 receptor markedly protects against rheumatoid-arthritis-associated inflammation and cartilage destruction in *A20^{myel-KO}* mice. These results reveal A20 as a novel negative regulator of Nlrp3 inflammasome activation, and describe *A20^{myel-KO}* mice as the first experimental model to study the role of inflammasomes in the pathology of rheumatoid arthritis.

A20 was deleted in myeloid cells by crossing *A20^{fllox/fllox}* mice into lyszyme M (LysM)-Cre-recombinase-expressing mice. Unlike wild-type macrophages, *A20^{myel-KO}* macrophages failed to induce A20 messenger RNA (mRNA) and protein expression in response to the Toll-like receptor-4 (TLR4) ligand lipopolysaccharide (LPS) (Fig. 1a, b), demonstrating the effectiveness of LysM-driven deletion of A20 in myeloid cells. Arthritis development in *A20^{myel-KO}* mice was shown to be independent of tumour necrosis factor receptor 1 (TNF-R1), whereas deletion of MyD88 markedly protected against pathology of rheumatoid arthritis². As this signalling adaptor operates downstream of both TLRs and interleukin-1 receptor (IL-1R), we crossed *Il1r1^{-/-}* mice into *A20^{myel-KO}* mice to assess the contribution of IL-1 signalling to arthritis pathogenesis. As expected, wild-type mice (*A20^{fllox/fllox} Il1r1^{+/+}*) did not develop arthritis, whereas *A20^{myel-KO}* mice spontaneously developed an arthritic phenotype (Fig. 1c). The incidence of *A20^{myel-KO}* mice developing arthritis was 100% (Fig. 1d). In sharp contrast, *A20^{myel-KO} Il1r1^{-/-}* mice were virtually devoid of clinical signs of arthritis (Fig. 1c, d). In agreement, clinical scoring of disease severity confirmed *A20^{myel-KO} Il1r1^{+/+}* mice as developing severe arthritic disease (high clinical score) whereas *A20^{myel-KO} Il1r1^{-/-}* mice were markedly protected (clinical

score 0). *A20^{myel-KO}* littermates that were heterozygous for IL-1R1 expression showed an intermediate arthritic phenotype between those of *A20^{myel-KO} Il1r1^{+/+}* and *A20^{myel-KO} Il1r1^{-/-}* mice (Fig. 1d, e). These clinical assessments were supported by a histological examination of representative ankle joints. Tissue sections of diseased *A20^{myel-KO} Il1r1^{+/+}* mice stained by haematoxylin and eosin showed significant synovial and periarticular inflammation and high levels of infiltrated mononuclear cells, which was associated with extensive cartilage and bone destruction (Fig. 1f). In marked contrast, ankle joints of *A20^{myel-KO} Il1r1^{-/-}* littermates were strongly protected from arthritic histopathology and contained significantly reduced numbers of infiltrating inflammatory cells (Fig. 1f). Semi-quantitative scoring of these histological parameters confirmed that the severity of arthritis was substantially lower in *A20^{myel-KO} Il1r1^{-/-}* mice relative to *A20^{myel-KO} Il1r1^{+/+}* littermates (Fig. 1g and Extended Data Table 1). These results demonstrate that IL-1 production is detrimental for arthritis pathogenesis in mice with a myeloid cell-restricted deletion in A20.

Macrophages are a prime source of proIL-1 β , and generally depend on caspase-1 for maturation and secretion of the biologically active cytokine. Caspase-1 is produced as a cytosolic zymogen, the activation of which is controlled by different inflammasomes³. To study the role of A20 in inflammasome signalling, we assessed caspase-1 processing in bone-marrow-derived macrophages (BMDMs) of wild-type and *A20^{myel-KO}* mice. Notably, caspase-1 activation was substantially increased in LPS-primed *A20^{myel-KO}* macrophages that were treated with soluble (ATP and nigericin) or crystalline (silica) stimuli of the Nlrp3 inflammasome compared with wild-type BMDMs (Fig. 2a). Concurrently, the levels of secreted IL-1 β in the culture medium of ATP- and nigericin-treated *A20^{myel-KO}* macrophages were about twice those of wild-type cells, and silica triggered nearly four times higher levels of secreted IL-1 β (Fig. 2b). Enhanced caspase-1 autoprocessing (Fig. 2c, d) and IL-1 β secretion (Fig. 2e, f) by the Nlrp3 inflammasome was evident within 10 min after ATP or nigericin addition, and continued to increase in a time-dependent fashion. Similarly, the induction of caspase-1-dependent pyroptosis was enhanced in *A20^{myel-KO}* macrophages (Fig. 2g, h). Despite their hypersensitivity for Nlrp3 inflammasome activation, *A20^{myel-KO}* macrophages failed to process caspase-1 and secrete IL-1 β and IL-18 when treated with LPS, ATP or nigericin alone (Extended Data Fig. 1). The increased responsiveness of *A20^{myel-KO}* macrophages towards inflammasome activation was restricted to the Nlrp3 inflammasome because caspase-1 processing and pyroptotic cell death by the Nlrp4 inflammasome were similarly induced in wild-type and *A20^{myel-KO}* macrophages that had been infected with *Salmonella enterica* serovar Typhimurium (Fig. 2i, j). Similarly, stimulation of the AIM2 inflammasome with cytosolic double-stranded DNA (dsDNA) did not result in differential levels of caspase-1 processing and pyroptosis induction in wild-type and *A20^{myel-KO}* macrophages (Fig. 2k, l). It is worth noting, however, that despite normal caspase-1 activation and pyroptosis levels in response

¹Department of Medical Protein Research, VIB, Ghent B-9000, Belgium. ²Department of Biochemistry, Ghent University, Ghent B-9000, Belgium. ³Department of Rheumatology, Ghent University, Ghent B-9000, Belgium. ⁴Department of Immunology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ⁵Inflammation Research Center, VIB, Ghent B-9052, Belgium. ⁶Department of Biomedical Molecular Biology, Ghent University, Ghent B-9052, Belgium.

*These authors contributed equally to this work.

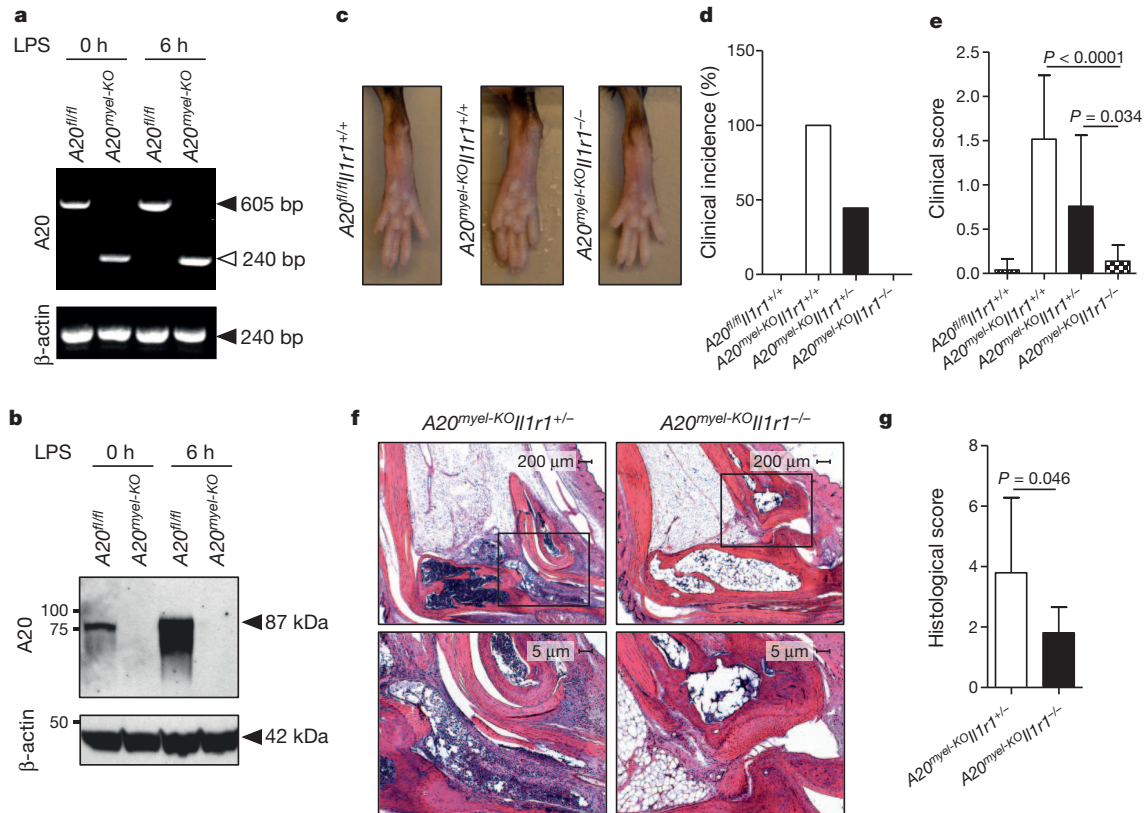


Figure 1 | *Il1r1* deficiency rescues the arthritis phenotype of $A20^{myel-KO}$ mice. **a**, **b**, A20 and β -actin mRNA (**a**) and protein (**b**) levels of LPS-stimulated BMDMs. **c**, Hind paws of 20-week-old mice. **d**, **e**, $A20^{fl/fl}Il1r1^{+/+}$ ($n = 19$), $A20^{myel-KO}Il1r1^{+/+}$ ($n = 8$), $A20^{myel-KO}Il1r1^{+/+}$ ($n = 15$) and $A20^{myel-KO}Il1r1^{-/-}$ ($n = 9$) mice aged 21–30 weeks were clinically scored for arthritis

to *S. enterica* serovar Typhimurium infection and dsDNA transfection, secretion of IL-1 β in response to these treatments was consistently higher in $A20^{myel-KO}$ macrophages compared with wild-type controls (Fig. 2m, n), which could be explained by increased induction of proIL-1 β mRNA in $A20^{myel-KO}$ macrophages (data not shown). Together, these results demonstrate that A20 negatively regulates activation of caspase-1 by the Nlrp3 inflammasome, but not by the Nlr4 and AIM2 inflammasomes.

Activation of the Nlrp3 inflammasome in wild-type macrophages is tightly regulated at different levels. A priming signal (referred to as step 1 and usually provided by TLRs) upregulates Nlrp3 expression levels along with the inflammasome substrate proIL-1 β via the pro-inflammatory transcription factor NF- κ B⁴. A20 negatively regulates LPS-induced NF- κ B activation (refs 5–8 and Extended Data Fig. 2a), which was also reflected in increased secretion of the NF- κ B-dependent cytokines IL-6 and TNF in $A20^{myel-KO}$ macrophages (Extended Data Fig. 2b, c). We further showed A20-deficiency to markedly enhance LPS-induced mRNA expression levels of Nlrp3 (Fig. 3a) and proIL-1 β (Fig. 3b). In contrast, LPS-induced transcript levels of caspase-1 and the inflammasome adaptor ASC were respectively mildly upregulated and normal in $A20^{myel-KO}$ macrophages (Extended Data Fig. 3a, b). Analysis of protein expression levels confirmed Nlrp3 and proIL-1 β to be significantly higher in LPS-primed $A20^{myel-KO}$ macrophages than wild-type cells, whereas caspase-1 and ASC were not differentially modulated in the two genotypes (Fig. 3c).

TLR stimulation during brief time intervals (10 min or less) was recently shown to license activation of the Nlrp3 inflammasome independently of new protein synthesis^{9–12}. Rapid Nlrp3 inflammasome activation resulted in procaspase-1 processing and secretion of pre-synthesized proIL-18 in the absence of the NF- κ B-dependent cytokines IL-1 β , TNF and IL-6 (refs 9–11). To address whether A20 modulated rapid Nlrp3 inflammasome activation, cells were exposed to ATP or nigericin after being primed with LPS for 10 min. As reported^{9–12}, wild-type BMDMs

incidence (**d**) and severity (**e**). **f**, Ankle joint sections stained with haematoxylin and eosin; magnification $\times 40$ (top) and $\times 100$ (bottom). **g**, Histological scores of ankle sections of $A20^{myel-KO}Il1r1^{+/+}$ ($n = 10$) and $A20^{myel-KO}Il1r1^{-/-}$ ($n = 8$). P values in **e** and **g** were determined by Student's *t*-test.

activated caspase-1 (Fig. 3d) and secreted significant amounts of IL-18, but not IL-1 β , TNF or IL-6 (Fig. 3e). Moreover, we noted Nlrp3 protein levels were lowered after stimulation both in wild-type and $A20^{myel-KO}$ macrophages (Fig. 3d), supporting the notion that acute Nlrp3 inflammasome activation occurred independently of LPS-induced new protein synthesis. Both caspase-1 processing and IL-18 secretion were markedly increased in $A20^{myel-KO}$ macrophages in the absence of substantial IL-1 β , TNF and IL-6 secretion (Fig. 3d, e). This was probably due to increased basal expression of Nlrp3 and proIL-18 in these cells (Fig. 3c, d). In agreement, basal mRNA levels of Nlrp3, proIL-1 β and proIL-18 were increased in untreated $A20^{myel-KO}$ macrophages (Fig. 3a, b and Extended Data Fig. 3c). Together, these results suggest that A20 negatively regulates Nlrp3 inflammasome signalling by suppressing NF- κ B-dependent production of Nlrp3 and the inflammasome substrates proIL-1 β and proIL-18. In agreement, the pharmacological inhibitor of kappa B kinase (IKK) inhibitor BMS-345541 significantly reduced Nlrp3 levels in LPS-primed $A20^{myel-KO}$ macrophages (Fig. 3f). Moreover, both BMS-345541 and the selective IKK2 inhibitor TPCA-1 significantly reduced ATP- and nigericin-induced caspase-1 autoprocessing, IL-1 β secretion and pyroptosis induction in LPS-primed $A20^{myel-KO}$ macrophages (Fig. 3g–i).

Having established A20 as a negative regulator of Nlrp3 inflammasome activation, we hypothesized that excessive Nlrp3 activation might drive pathology of rheumatoid arthritis in A20-deficient mice upstream of IL-1R1. To test this hypothesis, *Nlrp3*^{-/-} mice were crossed into $A20^{myel-KO}$ mice and the levels of four cytokines (IL-1 α , IL-1 β , IL-6 and TNF) relevant to rheumatoid arthritis were monitored. Although IL-1 α levels were not significantly different in A20-sufficient and $A20^{myel-KO}$ mice (Extended Data Fig. 4a), the latter group of mice had significantly higher levels of IL-1 β in circulation (Fig. 4a). In addition, the levels of IL-6 and TNF were also significantly higher in $A20^{myel-KO}$ mice compared with $A20^{flox/flox}$ littermates (Extended Data Fig. 4b, c). Notably,

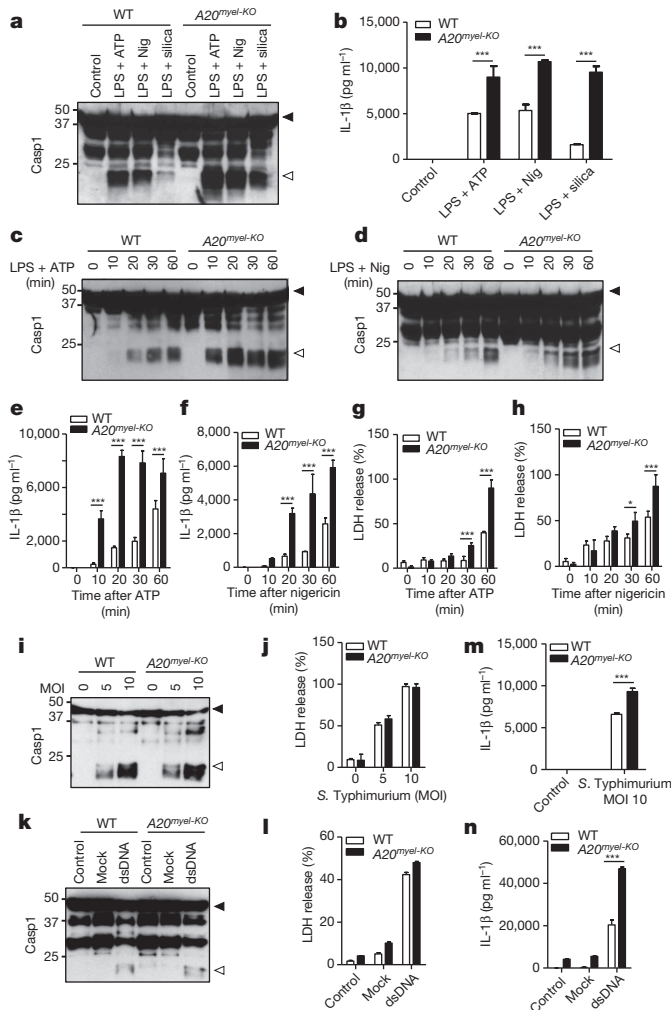


Figure 2 | Hyperactivation of the Nlrp3 but not the Nlr4 and AIM2 inflammasomes, in A20-deficient macrophages. **a–h**, BMDMs were stimulated as described in Methods. Lysates were immunoblotted for caspase-1 (**a**, **c**, **d**) and supernatants analysed for IL-1 β (**b**, **e**, **f**) and LDH (**g**, **h**). Nig, nigericin. **i–n**, BMDMs were treated as described in Methods. Lysates were immunoblotted for caspase-1 (**i**, **k**) and supernatants analysed for LDH (**j**, **l**) and IL-1 β (**m**, **n**). Black arrow, procaspase-1; white arrow, p20. MOI, multiplicity of infection. Data represent mean \pm s.d. of one out of three biological replicates, with three technical replicates each ($^*P < 0.05$; $^{***}P < 0.001$; Student's *t*-test).

deletion of Nlrp3 in *A20^{myel-KO}* mice markedly reduced IL-1 β secretion to baseline levels of *A20^{flox/flox}* mice, thereby demonstrating that the Nlrp3 inflammasome contributes critically to excessive IL-1 β production in *A20^{myel-KO}* mice *in vivo* (Fig. 4a). Intriguingly, *A20^{myel-KO} Nlrp3^{-/-}* mice also were protected from excessive IL-6 production, suggesting that high IL-6 levels are consequent to excessive inflammasome-mediated IL-1 β production (Extended Data Fig. 4b). In contrast, TNF production was not significantly affected in *A20^{myel-KO} Nlrp3^{-/-}* mice (Extended Data Fig. 4c). In agreement, TNF-R1 signalling was previously shown to be dispensable for pathology of rheumatoid arthritis in *A20^{myel-KO}* mice, whereas IL-6 neutralization provided protection².

Based on these findings, we assessed the contribution of Nlrp3 to pathogenesis of rheumatoid arthritis in *A20^{myel-KO}* mice. Swelling and redness of the hind paws of *A20^{myel-KO} Nlrp3^{+/+}* mice became evident around 11 weeks of age (Fig. 4b), and had afflicted all animals of this genotype when they were 20 weeks old (Fig. 4c). Disease severity continued to progress, and became increasingly pronounced in *A20^{myel-KO} Nlrp3^{+/+}* mice aged 21–40 weeks (Fig. 4d). In contrast, *A20^{myel-KO} Nlrp3^{-/-}* mice of similar age were markedly protected from rheumatoid arthritis,

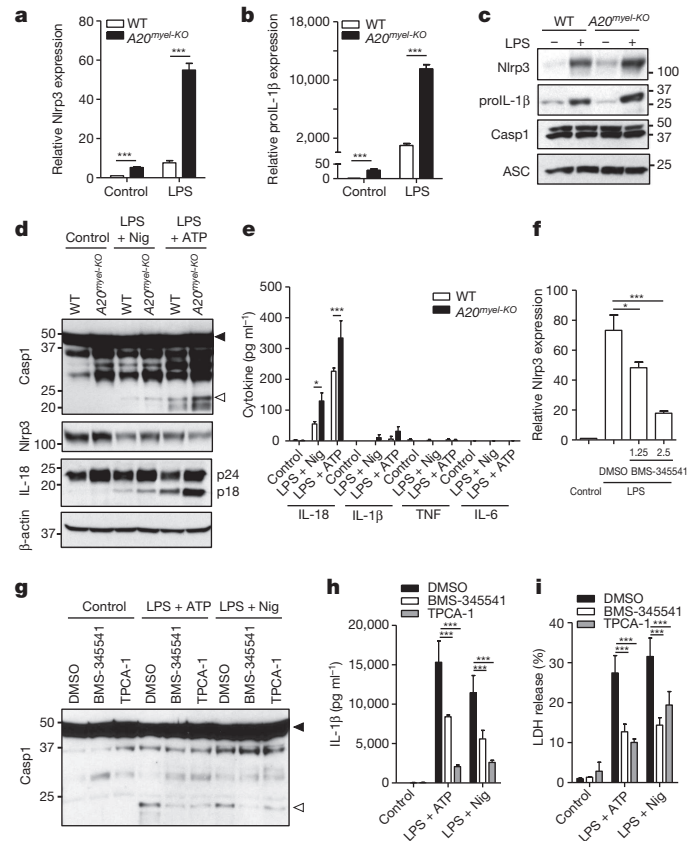


Figure 3 | A20 inhibits Nlrp3 inflammasome priming. **a**, **b**, Nlrp3 (**a**) and proIL-1 β (**b**) mRNA levels of LPS-treated BMDMs. **c**, Expression of the indicated proteins in BMDMs 6 h after LPS treatment. **d**, **e**, Rapid Nlrp3 inflammasome activation as described in Methods. Expression of the indicated proteins (**d**) and secreted cytokines (**e**) were determined. **f–i**, *A20^{myel-KO}* BMDMs were treated as indicated. Nlrp3 mRNA levels (**f**), caspase-1 expression (**g**), secreted IL-1 β (**h**) and LDH activity (**i**) were determined. Black arrow, procaspase-1; white arrow, p20. Data represent mean \pm s.d. of one out of three biological replicates, with three technical replicates each ($^*P < 0.05$; $^{***}P < 0.001$; Student's *t*-test).

and their hind paws had a normal appearance and lacked clinical signs of pathology of rheumatoid arthritis (Fig. 4b–d). Histological analysis of the ankle joints of these mice showed significantly reduced synovial and periarticular inflammation, and substantially less infiltrated mononuclear cells compared with tissue sections of *A20^{myel-KO} Nlrp3^{+/+}* mice of comparable age (Fig. 4e, f and Extended Data Table 2). In agreement, three-dimensional micro-computed tomography imaging showed that the extent of bone erosion in hind paws of representative *A20^{myel-KO} Nlrp3^{+/+}* and *A20^{myel-KO} Nlrp3^{-/-}* mice was markedly different. Unlike in Nlrp3-deficient *A20^{myel-KO}* mice, hind paws of Nlrp3-sufficient mice exhibited severe loss of bone density in the metatarsal region (Fig. 4g), demonstrating a key role for Nlrp3 in the pathology of rheumatoid arthritis.

We also analysed the impact of caspase-1/11 deficiency on IL-1 β secretion and pathology of rheumatoid arthritis in *A20^{myel-KO}* mice. IL-1 β levels in circulation were significantly reduced in *A20^{myel-KO} Casp1/11^{-/-}* mice compared with *A20^{myel-KO} Casp1/11^{+/+}* mice (Fig. 4h). As in *A20^{myel-KO} Nlrp3^{-/-}* mice, serum levels of IL-6 were markedly reduced in *A20^{myel-KO} Casp1/11^{-/-}* mice, whereas TNF production was not significantly different compared with *A20^{myel-KO} Casp1/11^{+/+}* mice (Extended Data Fig. 4). Moreover, hind paws of all analysed *A20^{myel-KO} Casp1/11^{+/+}* mice were clearly inflamed and swollen (Fig. 4i–k). In contrast, 50% of *A20^{myel-KO} Casp1/11^{-/-}* mice were devoid of clinical signs of arthritis, and disease symptoms in the remaining *A20^{myel-KO} Casp1/11^{-/-}* mice were very mild (Fig. 4i–k). In agreement, analysis of

14. Kolly, L. *et al.* Inflammatory role of ASC in antigen-induced arthritis is independent of caspase-1, NALP-3, and IPAF. *J. Immunol.* **183**, 4003–4012 (2009).
15. Consortium, W. T. C. C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
16. Dieguez-Gonzalez, R. *et al.* Analysis of TNFAIP3, a feedback inhibitor of nuclear factor- κ B and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility. *Arthritis Res. Ther.* **11**, R42 (2009).
17. Plenge, R. M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nature Genet.* **39**, 1477–1482 (2007).
18. Thomson, W. *et al.* Rheumatoid arthritis association at 6q23. *Nature Genet.* **39**, 1431–1433 (2007).
19. Ben Hamad, M. *et al.* Association study of CARD8 (p.C10X) and NLRP3 (p.Q705K) variants with rheumatoid arthritis in French and Tunisian populations. *Int. J. Immunogenet.* **39**, 131–136 (2012).
20. Kastbom, A. *et al.* Genetic variation in proteins of the cryopyrin inflammasome influences susceptibility and severity of rheumatoid arthritis (the Swedish TIRA project). *Rheumatology (Oxford)* **47**, 415–417 (2008).
21. Mathews, R. J. *et al.* Evidence of NLRP3-inflammasome activation in rheumatoid arthritis (RA); genetic variants within the NLRP3-inflammasome complex in relation to susceptibility to RA and response to anti-TNF treatment. *Ann. Rheum. Dis.* **73**, 1202–1210 (2014).
22. Mariathasan, S. *et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
23. Schott, W. H. *et al.* Caspase-1 is not required for type 1 diabetes in the NOD mouse. *Diabetes* **53**, 99–104 (2004).
24. Glaccum, M. B. *et al.* Phenotypic and functional characterization of mice that lack the type I receptor for IL-1. *J. Immunol.* **159**, 3364–3371 (1997).

Acknowledgements We thank R. Flavell and V. Dixit for supplying mutant mice. L.V.W. is a postdoctoral fellow with the Fund for Scientific Research-Flanders (FWO). This work was supported by the Ghent University Concerted Research Actions (grant BOF14/GOA/013) and grants from the European Research Council (Grant 281600) and the FWO (grants G030212N, 1.2.201.10.N.00 and 1.5.122.11.N.00) to M.L., and by FWO research grants (Odysseus-G091908, G061910N and G016812N) and a Queen Elisabeth Medical Foundation grant to G.V.L., and by the Ghent University Group-ID MRP to G.V.L., R.B. and D.E. T.-D.K. is supported by grants from the National Institutes of Health (AR056296, CA163507 and AI101935) and the American Lebanese Syrian Associated Charities.

Author Contributions L.V.W., G.V.L. and M.L. designed the study; L.V.W., N.V., A.F., P.J., E.V. and P.V. performed experiments; L.V.W., N.V., P.J., P.V., G.V.L., R.B., D.E., T.-D.K. and M.L. analysed data and wrote the manuscript; T.-D.K. provided essential reagents and scientific insight; M.L. oversaw the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.L. (mohamed.lamkanfi@vib-ugent.be).

Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys

James B. Whitney^{1,2}, Alison L. Hill³, Srisowmya Sanisetty¹, Pablo Penaloza-MacMaster¹, Jinyan Liu¹, Mayuri Shetty¹, Lily Parenteau¹, Crystal Cabral¹, Jennifer Shields¹, Stephen Blackmore¹, Jeffrey Y. Smith¹, Amanda L. Brinkman¹, Lauren E. Peter¹, Sheeba I. Mathew¹, Kaitlin M. Smith¹, Erica N. Borducchi¹, Daniel I. S. Rosenbloom³, Mark G. Lewis⁴, Jillian Hattersley⁵, Bei Li⁵, Joseph Hesselgesser⁵, Romas Geleziunas⁵, Merlin L. Robb⁶, Jerome H. Kim⁶, Nelson L. Michael⁶ & Dan H. Barouch^{1,2}

The viral reservoir represents a critical challenge for human immunodeficiency virus type 1 (HIV-1) eradication strategies^{1–5}. However, it remains unclear when and where the viral reservoir is seeded during acute infection and the extent to which it is susceptible to early antiretroviral therapy (ART). Here we show that the viral reservoir is seeded rapidly after mucosal simian immunodeficiency virus (SIV) infection of rhesus monkeys and before systemic viraemia. We initiated suppressive ART in groups of monkeys on days 3, 7, 10 and 14 after intrarectal SIV_{MAC251} infection. Treatment with ART on day 3 blocked the emergence of viral RNA and proviral DNA in peripheral blood and also substantially reduced levels of proviral DNA in lymph nodes and gastrointestinal mucosa as compared with treatment at later time points. In addition, treatment on day 3 abrogated the induction of SIV-specific humoral and cellular immune responses. Nevertheless, after discontinuation of ART following 24 weeks of fully suppressive therapy, virus rebounded in all animals, although the monkeys that were treated on day 3 exhibited a delayed viral rebound as compared with those treated on days 7, 10 and 14. The time to viral rebound correlated with total viraemia during acute infection and with proviral DNA at the time of ART discontinuation. These data demonstrate that the viral reservoir is seeded rapidly after intrarectal SIV infection of rhesus monkeys, during the ‘eclipse’ phase, and before detectable viraemia. This strikingly early seeding of the refractory viral reservoir raises important new challenges for HIV-1 eradication strategies.

The viral reservoir in memory CD4⁺ T cells in HIV-1-infected individuals cannot be eliminated by current antiretroviral drugs or HIV-1-specific immune responses^{1–5}. This archive of replication-competent virus is the source of viral rebound in nearly all HIV-1-infected individuals who discontinue ART^{3,5} and represents a critical hurdle for HIV-1 eradication strategies^{6,7}. The temporal dynamics of seeding the viral reservoir have not been previously defined but have been presumed to occur during peak viraemia in acute HIV-1 infection. To evaluate the impact of early ART on the viral reservoir, we initiated suppressive ART at various time points after mucosal SIV infection of rhesus monkeys.

We inoculated 20 Indian-origin adult rhesus monkeys (*Macaca mulatta*) that did not express the protective major histocompatibility complex (MHC) class I alleles *Mamu-A*01*, *Mamu-B*08* and *Mamu-B*17* with 500 median tissue culture infective doses (TCID₅₀) of SIV_{MAC251} (refs 8–10) by the intrarectal route. We initiated ART on days 3, 7, 10 and 14 after infection with a pre-formulated cocktail of tenofovir, emtricitabine and dolutegravir (see Methods), and a control group received no ART (*n* = 4 per group). ART was administered daily by subcutaneous injection for 24 weeks. Treatment on day 3 after infection resulted in no detectable viraemia (<50 RNA copies ml⁻¹)¹¹ at any time point in four of four monkeys (Fig. 1a). In contrast, treatment on days 7, 10 and 14 abruptly interrupted the exponential growth of the virus and reduced plasma viral RNA to undetectable levels within 3–4 weeks. The mean levels of plasma viral

RNA at the time of ART initiation in these groups of monkeys were 5.88 log copies ml⁻¹ (day 7), 7.11 log copies ml⁻¹ (day 10) and 7.50 log copies ml⁻¹ (day 14), which were comparable with the levels of plasma viral RNA in untreated controls at these time points (Fig. 1b). Viral dynamics modelling¹² revealed an initial exponential growth rate of 1.5 ± 0.5 per day, corresponding to a basic reproductive ratio of $R_0 = 9.5 \pm 5.1$ (see Methods; Extended Data Fig. 1 and Extended Data Table 1). An exponential decay rate of plasma viraemia after ART initiation of 0.60 ± 0.17 per day was observed in all the treated groups, corresponding to a 1.3 ± 0.4 day half-life of infected cells (Extended Data Fig. 1).

After initial control of viraemia, all animals treated with ART exhibited undetectable plasma viral loads (<50 RNA copies ml⁻¹) for the full

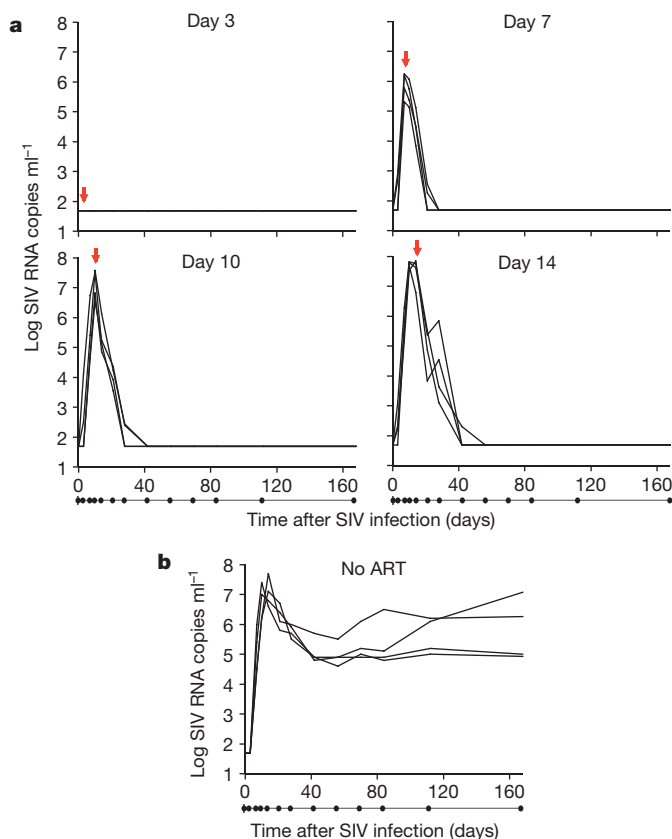


Figure 1 | Viral decay kinetics after treatment with ART. **a**, **b**, Log plasma viral RNA (copies ml⁻¹) in rhesus monkeys infected with SIV_{MAC251} and after initiation of ART on days 3, 7, 10 and 14 of infection (**a**) or with no ART (**b**). Assay sensitivity is 50 RNA copies ml⁻¹. Red arrows indicate initiation of ART. Black dots below x-axis indicate sampling time points.

¹Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts 02139, USA. ³Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts 02138 USA. ⁴Bioqual, Rockville, Maryland 20852, USA. ⁵Gilead Sciences, Foster City, California 94404, USA. ⁶US Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, Maryland 20910, USA.

24-week course of suppressive therapy with no detectable viral blips (Fig. 1a), demonstrating the potency and consistency of this ART regimen. Moreover, ultrasensitive plasma viral load assays at week 20 also proved negative (≤ 6 RNA copies ml^{-1})¹³ in all animals (Extended Data Fig. 2). In addition, viral sequences from stimulated peripheral blood mononuclear cells (PBMCs) from ART-suppressed, SIV-infected monkeys using the same ART regimen revealed no viral sequence evolution over 6 months in a separate study (J.B.W., unpublished observations). Furthermore, treatment intensification studies in other SIV-infected rhesus monkeys in which the protease inhibitor darunavir was added to the current ART regimen did not lead to enhanced virological control (R.G., unpublished observations). Taken together, these data indicate that the ART regimen that was used in the present study was fully suppressive.

We next assessed the development of SIV-specific humoral and cellular immune responses in these animals. Monkeys treated on day 3 after infection developed no detectable SIV Env-specific antibody responses by enzyme-linked immunosorbent assay (ELISA) (Fig. 2a) and no detectable SIV Env-, Pol- or Gag-specific T lymphocyte responses by interferon (IFN)- γ ELISPOT assays (Fig. 2b) at weeks 4, 10, and 20 or 24 of infection. In contrast, monkeys treated on days 7, 10 and 14 developed detectable but lower SIV-specific humoral and cellular immune responses as compared with untreated controls, presumably as a result of reduced antigenic stimulus after ART initiation. Multiparameter intracellular cytokine staining (ICS) assays confirmed that SIV Gag-specific CD8⁺ and CD4⁺ T lymphocyte responses were undetectable in animals treated on day 3 and were lower in animals treated on days 7, 10 and 14 as compared with untreated controls (Extended Data Figs 3 and 4). Gag-specific CD8⁺ and CD4⁺ T lymphocytes in monkeys treated on days 7, 10 and 14 also

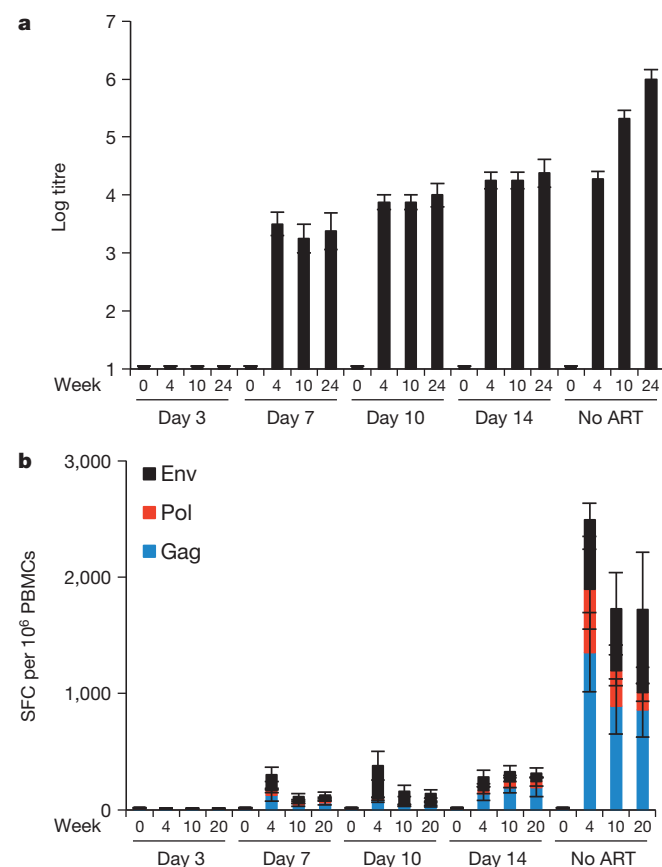


Figure 2 | SIV-specific humoral and cellular immune responses during ART. **a**, **b**, Env-specific ELISA antibody titres at weeks 0, 4, 10 and 24 (**a**) and Env-, Pol-, and Gag-specific IFN- γ ELISPOT responses at weeks 0, 4, 10 and 20 in SIV-infected monkeys that initiated ART on days 3, 7, 10 and 14 of infection or with no ART (**b**). Mean responses are shown ($N = 4$ animals per group). SFC, spot-forming cells. Error bars show standard error of the mean (s.e.m.).

exhibited reduced immune activation and proliferation as measured by Ki67 expression (Extended Data Fig. 4). These data demonstrate that initiation of ART on day 3 blocked the emergence of plasma viraemia and abrogated the induction of SIV-specific humoral and cellular immune responses.

We next determined the impact of early ART on levels of proviral DNA¹⁴ in PBMCs, lymph node mononuclear cells (LNCs) and gastro-intestinal mucosa mononuclear cells (GMMCs) over the course of 24 weeks of treatment with suppressive ART. In monkeys that initiated ART on day 3, there was a striking anatomical discordance with no proviral DNA detected in PBMCs at any time point (< 3 DNA copies per 10^6 cells) (Fig. 3a). In contrast, clear but low levels of proviral DNA were detected in inguinal LNCs and in colorectal GMMCs in these animals, although proviral DNA declined to undetectable or nearly undetectable levels in three of four of these animals by week 24. In monkeys treated with ART on days 7, 10 and 14, proviral DNA was readily detected in PBMCs as well as in LNCs and GMMCs (Fig. 3b–d). Moreover, in animals treated with ART on days 10 and 14, proviral DNA in LNCs appeared to stabilize by week 12 with minimal subsequent decline, consistent with a stable viral reservoir (Fig. 3c, d). In untreated animals, proviral DNA was markedly higher than in ART-treated animals with minimal decline over time (Fig. 3e). Analysis of sorted cell subpopulations demonstrated that proviral DNA was found primarily in central memory and transitional memory CD4⁺ T lymphocytes in lymph nodes on day 3 and in both PBMCs and lymph nodes on day 7 after SIV infection (Extended Data Fig. 5).

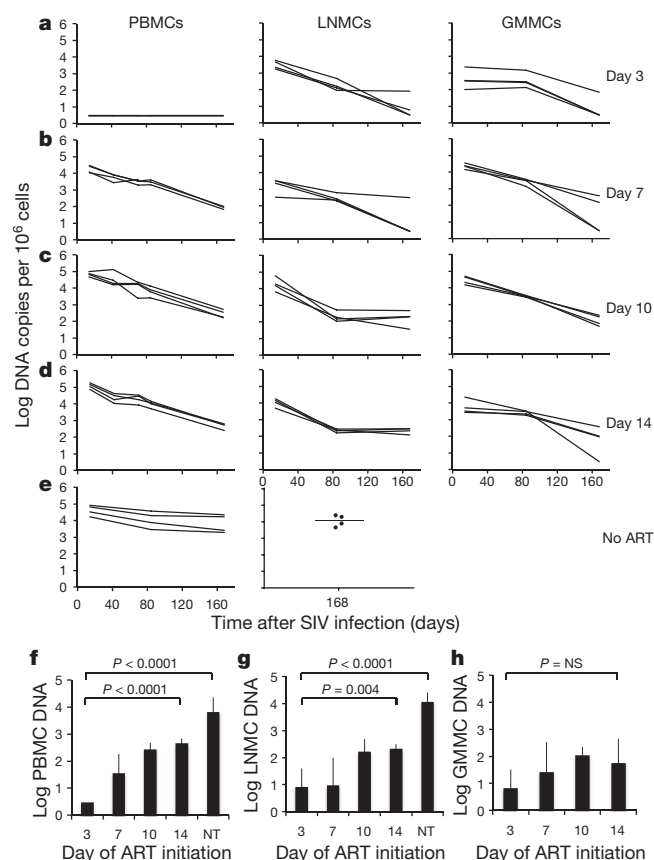


Figure 3 | Proviral DNA during ART. **a–e**, Log proviral DNA (copies per 10^6 cells) in PBMCs, LNCs and GMMCs in monkeys that initiated ART on days 3 (**a**), 7 (**b**), 10 (**c**) and 14 (**d**) of infection or with no ART (**e**). **f–h**, Comparisons of mean levels of log proviral DNA per 10^6 cells at the time of ART discontinuation (week 24) in ART-treated and untreated (NT) monkeys in PBMCs (**f**), LNCs (**g**) and GMMCs (**h**) are also shown ($N = 4$ animals per group). Assay sensitivity is 3 DNA copies per 10^6 cells. P values reflect one-sided t -tests. NS, not significant. Error bars show standard deviation.

These data indicate that initiation of ART on day 3 reduced levels of proviral DNA at week 24 by at least 2.2 log in PBMCs ($P < 0.0001$; Fig. 3f), 1.0 log in LNMCs ($P = 0.004$; Fig. 3g), and 0.9 log in GMMCs ($P = \text{not significant}$; Fig. 3h) as compared with initiation of ART on day 14. The high variability in the GMMC samples probably reflected sampling variation in this anatomical compartment. Compared to untreated animals, initiation of ART on day 3 reduced levels of proviral DNA at week 24 by at least 3.3 log in PBMCs ($P < 0.0001$; Fig. 3f) and 3.1 log in LNMCs ($P < 0.0001$; Fig. 3g).

At week 24, ART was discontinued, and monkeys were monitored twice weekly for evidence of viral rebound. Viral rebound, defined as plasma viral RNA > 50 copies ml^{-1} , occurred in all animals (Fig. 4a). In particular, viral rebound occurred in four of four animals that initiated ART on day 3, albeit with threefold delayed kinetics as compared with animals that initiated ART at later time points (median 21 days to viral rebound in the day 3 treated animals compared with 7 days in the day 14 treated animals; $P < 0.001$; Fig. 4b). The median log setpoint viral load after rebound, defined as viral loads on days 56–112 following ART discontinuation, was also 1.04 log RNA copies ml^{-1} lower for all the ART-treated animals as compared with untreated controls (4.59 log RNA copies ml^{-1} for all treatment groups combined versus 5.63 log RNA copies ml^{-1} for untreated monkeys; $P = 0.01$; Fig. 4c), suggesting that there is a benefit to early ART, although no significant differences were observed among setpoint viral loads in the day 3, 7, 10 and 14 treatment groups. Setpoint viral loads in the untreated animals were comparable with historical controls^{9,10}. Taken together, these data show that the persistent viral reservoir was seeded by day 3 of infection and led to viral rebound in all monkeys after ART discontinuation.

To gain mechanistic insight into the kinetics of the viral rebound, we used viral dynamics modelling^{15,16} (see Methods; Extended Data Fig. 6

and Extended Data Table 2). Initiation of ART on day 3 as compared with days 7, 10 and 14 resulted in lower modelled residual viral loads at the time of ART discontinuation ($P = 0.01$) and a trend towards a greater viral growth rate R_0 during viral rebound ($P = 0.06$), but no difference in post-rebound setpoint viral loads (Extended Data Fig. 7). Average R_0 during viral rebound was 4.2 ± 1.8 in the day 3 treated animals as compared with 2.3 ± 0.6 in the day 14 treated animals ($P = 0.05$; Extended Data Fig. 7), presumably reflecting the partially effective SIV-specific immune responses in the latter group (Fig. 2). Total plasma viraemia during acute infection was interpolated and calculated as the area under the curve for pre-ART viral loads (AUC VL) (Extended Data Fig. 8). Consistent with recent findings in acute HIV-1 infection in humans¹⁶, the AUC VL during acute infection correlated with levels of proviral DNA in PBMCs ($P < 0.0001$; Fig. 5a), LNMCs ($P = 0.005$; Fig. 5b) and GMMCs ($P = 0.04$; data not shown) at the time of ART discontinuation. Moreover, both the AUC VL ($P < 0.0001$; Fig. 5c) and proviral DNA in PBMCs at the time of ART discontinuation ($P = 0.003$; Fig. 5d) correlated inversely with the interpolated time to viral rebound. These data suggest that total plasma viraemia during acute infection and proviral DNA immediately before ART discontinuation may predict the time to viral recrudescence.

We show that the viral reservoir can be seeded substantially earlier than previously recognized. After intrarectal SIV infection of rhesus monkeys, the viral reservoir was seeded during the first few days of infection, during the eclipse phase, and before detectable viraemia, probably in the mucosal and lymphoid tissues that represent the first sites of viral replication¹⁷. Consistent with this finding, we observed proviral DNA in lymph nodes and in gastrointestinal mucosa but not in PBMCs in monkeys treated on day 3 after infection (Fig. 3a and Extended Data Fig. 5). The observation that the viral reservoir can be seeded before detectable viraemia suggests that substantial pathogenesis occurs in tissues in the first few days after mucosal virus exposure and prior to virus replication in peripheral blood, which has important implications for HIV-1 therapeutic and eradication strategies.

Our data are concordant with recent clinical studies that have demonstrated that early ART can reduce the size of the viral reservoir and delay or reduce viral rebound after ART discontinuation in humans^{18–22}. Our findings similarly show that early ART decreased proviral DNA in blood and tissues (Fig. 3f–h) and delayed and reduced viral rebound (Fig. 4b, c) after ART discontinuation in SIV-infected rhesus monkeys. Moreover,

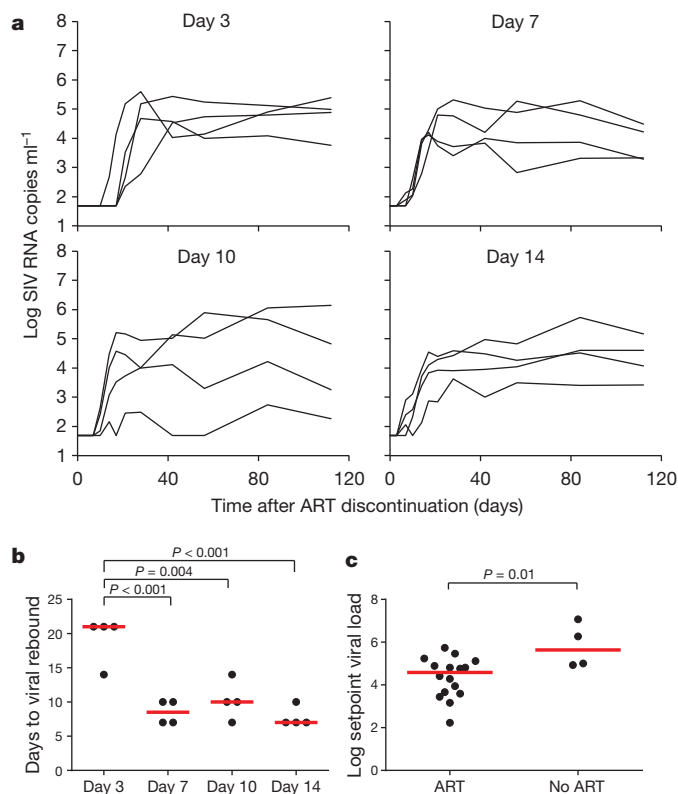


Figure 4 | Viral rebound kinetics after ART discontinuation. **a**, Log plasma viral RNA (copies ml^{-1}) in animals after discontinuation of ART at week 24 in monkeys that initiated ART on days 3, 7, 10 and 14 of infection. Assay sensitivity is 50 RNA copies ml^{-1} . **b**, Times to viral rebound, defined as the first time point at which plasma viral RNA was > 50 copies ml^{-1} . Red bars indicate median times to viral rebound. **c**, Setpoint viral loads following viral rebound, defined as day 56–112 after ART discontinuation compared with untreated monkeys. Red bars indicate median setpoint viral loads. P values reflect one-sided t -tests.

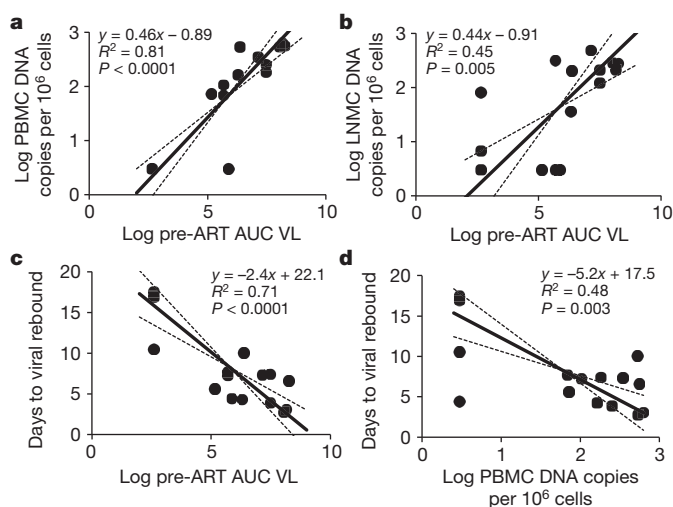


Figure 5 | Viral dynamics and correlations. **a**, **b**, Correlations of AUC VL with proviral DNA in PBMCs (**a**) and LNMCs (**b**) at the time of ART discontinuation are shown. **c**, **d**, Correlations of AUC VL (**c**) and proviral DNA in PBMCs before ART discontinuation (**d**) with the interpolated time to viral rebound are also shown. R^2 and P values were calculated from correlation analyses, and trend lines were calculated using standardized major axis regression. Animals with undetectable proviral DNA were plotted at the detection limit of 3 DNA copies (0.48 log) per 10^6 cells.

our observations extend previous studies that have shown effective post-exposure prophylaxis with short courses of ART when initiated 24 h after SIV infection in monkeys^{23,24}. However, in the present study, initiation of suppressive ART even as early as day 3 after infection failed to eliminate the viral reservoir and did not prevent viral rebound despite 24 weeks of fully suppressive ART. In addition, our data (Fig. 4a) are consistent with clinical studies that have shown that the vast majority of HIV-1-infected individuals who initiate ART during acute infection show viral rebound after discontinuation of ART^{25–27}.

Our findings contrast with the sustained remission and potential cure of a viraemic baby that was treated with ART at 30 h of life²⁸. It is possible that this baby was inoculated parenterally with maternal cells instead of mucosally with virus, resulting in rapid viraemia without a previraemic eclipse phase of viral replication in mucosal and lymphoid tissues. The positive outcome in this baby therefore might have reflected the route of transmission, the lack of an eclipse phase, the very rapid initiation of ART, and/or the paucity of memory CD4⁺ T lymphocytes in the neonatal immune system²⁹.

Clinical studies are required to confirm our observations, since important differences exist between SIV infection of rhesus monkeys and HIV-1 infection of humans. For example, the SIV dose used in the present study in monkeys was selected to limit the number of transmitted/founder viruses but also to infect all animals⁸ and thus was substantially higher than typical HIV-1 doses in humans. Nevertheless, the higher challenge dose has been shown to shorten the eclipse period and to lead to earlier plasma viraemia⁸. Additional differences may also exist between SIV-infected rhesus monkeys and HIV-1-infected humans.

The strikingly early seeding of the viral reservoir within the first few days of infection is sobering and presents new challenges to HIV-1 eradication strategies. If HIV-1 similarly seeds a persistent viral reservoir in mucosal and lymphoid tissues during the eclipse phase of infection and before viraemia after sexual exposure in humans, then it will be very difficult to initiate ART before reservoir seeding, since viraemia is typically used for the clinical diagnosis of acute HIV-1 infection. Taken together, our data suggest that extremely early initiation of ART, extended ART duration, and probably additional interventions that target the viral reservoir will be required for HIV-1 eradication. Moreover, an improved understanding of the virological parameters that predict viral rebound after ART discontinuation will help guide future HIV-1 eradication efforts. *Note added in proof:* Since the acceptance of this paper, the 'Mississippi baby' that was treated with ART at 30 h of life and had a prolonged remission off therapy has now developed detectable levels of HIV-1 replication and has restarted therapy, demonstrating that early ART did not eradicate the viral reservoir in this child.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 May; accepted 19 June 2014.

Published online 20 July 2014.

- Finzi, D. *et al.* Latent infection of CD4⁺ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nature Med.* **5**, 512–517 (1999).
- Zhang, L. *et al.* Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N. Engl. J. Med.* **340**, 1605–1613 (1999).
- Chun, T. W., Davey, R. T., Jr, Engel, D., Lane, H. C. & Fauci, A. S. Re-emergence of HIV after stopping therapy. *Nature* **401**, 874–875 (1999).
- Chun, T. W. *et al.* Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* **387**, 183–188 (1997).
- Chun, T. W. *et al.* Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl Acad. Sci. USA* **94**, 13193–13197 (1997).
- Persaud, D., Zhou, Y., Siliciano, J. M. & Siliciano, R. F. Latency in human immunodeficiency virus type 1 infection: no easy answers. *J. Virol.* **77**, 1659–1665 (2003).
- Ho, Y. C. *et al.* Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* **155**, 540–551 (2013).
- Liu, J. *et al.* Low-dose mucosal simian immunodeficiency virus infection restricts early replication kinetics and transmitted virus variants in rhesus monkeys. *J. Virol.* **84**, 10406–10412 (2010).

- Barouch, D. H. *et al.* Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* **482**, 89–93 (2012).
- Liu, J. *et al.* Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys. *Nature* **457**, 87–91 (2009).
- Cline, A. N., Bess, J. W., Piatak, M., Jr & Lifson, J. D. Highly sensitive SIV plasma viral load assay: practical considerations, realistic performance expectations, and application to reverse engineering of vaccines for AIDS. *J. Med. Primatol.* **34**, 303–312 (2005).
- Nowak, M. A. *et al.* Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection. *J. Virol.* **71**, 7518–7525 (1997).
- Palmer, S. *et al.* New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **41**, 4531–4536 (2003).
- Whitney, J. B. *et al.* T-cell vaccination reduces simian immunodeficiency virus levels in semen. *J. Virol.* **83**, 10840–10843 (2009).
- Rosenbloom, D. I., Hill, A. L., Rabi, S. A., Siliciano, R. F. & Nowak, M. A. Antiretroviral dynamics determines HIV evolution and predicts therapy outcome. *Nature Med.* **18**, 1378–1385 (2012).
- Archin, N. M. *et al.* Immediate antiviral therapy appears to restrict resting CD4⁺ cell HIV-1 infection without accelerating the decay of latent infection. *Proc. Natl Acad. Sci. USA* **109**, 9523–9528 (2012).
- Haase, A. T. Targeting early infection to prevent HIV-1 mucosal transmission. *Nature* **464**, 217–223 (2010).
- Ananworanich, J. *et al.* Impact of multi-targeted antiretroviral treatment on gut T cell depletion and HIV reservoir seeding during acute HIV infection. *PLoS ONE* **7**, e33948 (2012).
- von Wyl, V. *et al.* Early antiretroviral therapy during primary HIV-1 infection results in a transient reduction of the viral setpoint upon treatment interruption. *PLoS ONE* **6**, e27463 (2011).
- Hocqueloux, L. *et al.* Long-term antiretroviral therapy initiated during primary HIV-1 infection is key to achieving both low HIV reservoirs and normal T cell counts. *J. Antimicrob. Chemother.* **68**, 1169–1178 (2013).
- Sáez-Cirión, A. *et al.* Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog.* **9**, e1003211 (2013).
- Steingrover, R. *et al.* HIV-1 viral rebound dynamics after a single treatment interruption depends on time of initiation of highly active antiretroviral therapy. *AIDS* **22**, 1583–1588 (2008).
- Tsai, C. C. *et al.* Effectiveness of postinoculation (R)-9-(2-phosphorylmethoxypropyl) adenine treatment for prevention of persistent simian immunodeficiency virus SIV_{mac} infection depends critically on timing of initiation and duration of treatment. *J. Virol.* **72**, 4265–4273 (1998).
- Tsai, C. C. *et al.* Prevention of SIV infection in macaques by (R)-9-(2-phosphorylmethoxypropyl)adenine. *Science* **270**, 1197–1199 (1995).
- Sáez-Cirión, A. *et al.* Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog.* **9**, e1003211 (2013).
- Stöhr, W. *et al.* Duration of HIV-1 viral suppression on cessation of antiretroviral therapy in primary infection correlates with time on therapy. *PLoS ONE* **8**, e78287 (2013).
- Rosenberg, E. S. *et al.* Safety and immunogenicity of therapeutic DNA vaccination in individuals treated with antiretroviral therapy during acute/early HIV-1 infection. *PLoS ONE* **5**, e10555 (2010).
- Persaud, D. *et al.* Absence of detectable HIV-1 viremia after treatment cessation in an infant. *N. Engl. J. Med.* **369**, 1828–1835 (2013).
- Liu, J., Li, H., Iampietro, M. J. & Barouch, D. H. Accelerated heterologous adenovirus prime-boost SIV vaccine in neonatal rhesus monkeys. *J. Virol.* **86**, 7829–7835 (2012).

Acknowledgements We thank M. Pensiero, M. Marovich, C. Dieffenbach, W. Wagner, C. Gittens, J. Valley-Ogunro, M. Nowak, R. Siliciano, D. Persaud, S. Deeks, N. Chomont, J. Ananworanich, L. Picker, F. Stephens, R. Hamel, K. Kelly and L. Dunne for advice, assistance and reagents. The SIV_{MAC239} peptides were obtained from the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program. We acknowledge support from the US Army Medical Research and Materiel Command and the US Military HIV Research Program through its cooperative agreement with the Henry M. Jackson Foundation for the Advancement of Military Medicine (W81XWH-07-2-0067, W81XWH-11-2-0174); the NIH (AI060354, AI078526, AI084794, AI095985, AI096040, AI100645); and the Ragon Institute of MGH, MIT and Harvard. The views expressed in this manuscript are those of the authors and do not represent the official views of the Department of the Army or the Department of Defense.

Author Contributions J.B.W., R.G., M.L.R., J.H.K., N.L.M. and D.H.B. designed the studies and interpreted the data. J.B.W. and S.S. led the virological assays. P.P.-M., J.L., M.S., L.P., C.C., J.S., S.B., J.Y.S., A.L.B., L.E.P., E.N.B. and K.M.S. led the study operations and the immunological assays. A.L.H. and D.J.S.R. led the mathematical modelling and statistical analysis. M.G.L. led the clinical care of the rhesus monkeys. B.L., J.H.a., J.H.i. and R.G. developed the antiretroviral drug cocktail. J.B.W. and D.H.B. wrote the paper with all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.B. (dbarouch@bidmc.harvard.edu).

Neuropathy of haematopoietic stem cell niche is essential for myeloproliferative neoplasms

Lorena Arranz¹, Abel Sánchez-Aguilera¹, Daniel Martín-Pérez¹, Joan Isern¹, Xavier Langa¹, Alexandar Tzankov², Pontus Lundberg², Sandra Muntión³, Yi-Shiuan Tzeng⁴, Dar-Ming Lai⁴, Jürg Schwaller², Radek C. Skoda² & Simón Méndez-Ferrer¹

Myeloproliferative neoplasms (MPNs) are diseases caused by mutations in the haematopoietic stem cell (HSC) compartment. Most MPN patients have a common acquired mutation of Janus kinase 2 (*JAK2*) gene in HSCs^{1–4} that renders this kinase constitutively active, leading to uncontrolled cell expansion. The bone marrow microenvironment might contribute to the clinical outcomes of this common event. We previously showed that bone marrow nestin⁺ mesenchymal stem cells (MSCs) innervated by sympathetic nerve fibres regulate normal HSCs^{5,6}. Here we demonstrate that abrogation of this regulatory circuit is essential for MPN pathogenesis. Sympathetic nerve fibres, supporting Schwann cells and nestin⁺ MSCs are consistently reduced in the bone marrow of MPN patients and mice expressing the human *JAK2(V617F)* mutation in HSCs. Unexpectedly, MSC reduction is not due to differentiation but is caused by bone marrow neural damage and Schwann cell death triggered by interleukin-1 β produced by mutant HSCs. In turn, *in vivo* depletion of nestin⁺ cells or their production of CXCL12 expanded mutant HSC number and accelerated MPN progression. In contrast,

administration of neuroprotective or sympathomimetic drugs prevented mutant HSC expansion. Treatment with β_3 -adrenergic agonists that restored the sympathetic regulation of nestin⁺ MSCs^{5,6} prevented the loss of these cells and blocked MPN progression by indirectly reducing the number of leukaemic stem cells. Our results demonstrate that mutant-HSC-driven niche damage critically contributes to disease manifestation in MPN and identify niche-forming MSCs and their neural regulation as promising therapeutic targets.

The stem cell niche has recently emerged as an oncogenic unit and an important element in regulating cancer stem cells, including HSCs^{7–11}. Most MPN patients who do not carry the *BCR-ABL* fusion have an acquired mutation in Janus kinase 2 (*JAK2(V617F)*) in HSCs that results in constitutive kinase activity, leading to uncontrolled expansion of HSCs and erythroid, megakaryocytic and myeloid progenitors^{1–4}. Somatic mutations in thrombopoietin receptor^{12,13} or calreticulin^{14,15} genes are found in some MPN patients and additional HSC mutations also affect disease progression^{16,17}. Changes in the HSC microenvironment might also contribute to MPN development, and expansion of bone marrow (BM) fibroblasts and bone-forming cells suggests the participation of MSCs.

We previously reported that mouse BM nestin⁺ MSCs are required to maintain HSCs⁵ and that human BM nestin⁺ cells can expand HSCs¹⁸. Here we found that, despite elevated BM blood-vessel density in MPN patients, nestin⁺ cell number and nestin messenger RNA expression were markedly reduced (Fig. 1a, b). This was reproduced in *Mx1-cre;JAK2(V617F)*

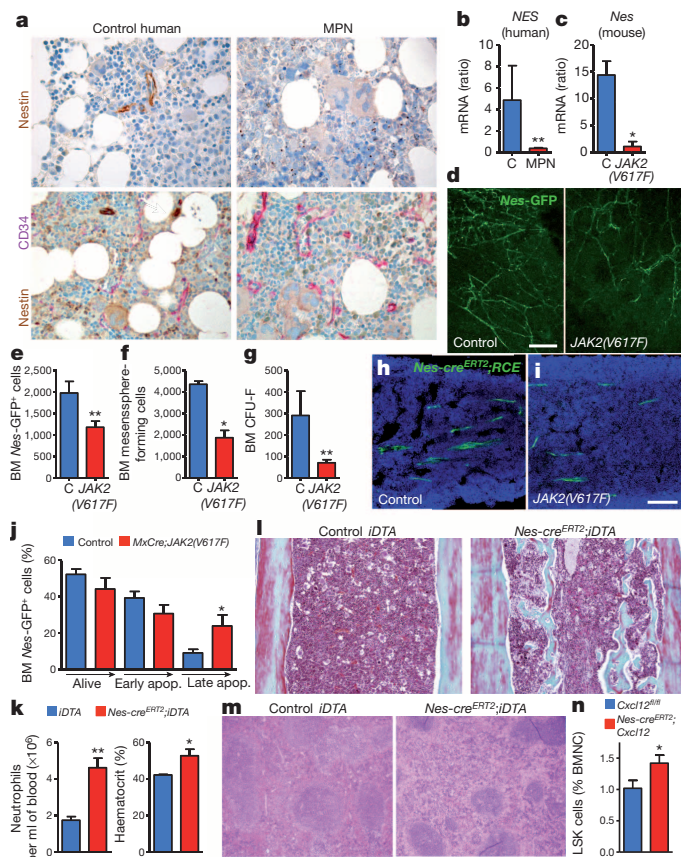


Figure 1 | Apoptosis of BM nestin⁺ HSC niche cells contributes to MPN progression. **a**, BM sections of controls (left) and MPN patients (right) immunostained with nestin (brown, all panels) and CD34 (red, lower panels; magnification $\times 200$). Numbers of nestin⁺ niches per mm² (mean \pm s.d.) were 1.15 ± 0.3 (control) and 0.17 ± 0.18 (MPN; $n = 40$; $P = 10^{-6}$, Mann-Whitney U test). **b**, **c**, Nestin mRNA expression in BM cells from controls ($n = 2$), MPN patients ($n = 11$) and mice ($n = 4$). **d**, Nes-GFP⁺ cells in skull BM of control (left) and MPN mice (right; $n = 10$). **e–g**, CD45⁺CD31⁺Ter119⁺Nes-GFP⁺ cells (**e**), clonal self-renewing mesenchymal sphere (mesosphere)-forming cells (**f**) and fibroblastic colony-forming units (CFU-F, **g**) in BM cells from wild-type mice 30 weeks after transplantation with control ($n = 7$) or MPN BM cells ($n = 12$). **h**, **i**, Lineage-tracing studies of nestin⁺ cells. Femoral sections of tamoxifen-treated Nes-cre^{ERT2};RCE:loxP mice 28 weeks after transplantation with control (**h**) or MPN BM cells (**i**), showing fluorescent signals from GFP and nuclei counterstained with DAPI ($n = 4$). **j**, Fraction of live, early and late apoptotic BM stromal Nes-GFP⁺ cells from control ($n = 7$) or MPN mice ($n = 5$). **k–m**, Blood counts (**k**), femoral trichrome (**l**) and spleen haematoxylin and eosin stainings (**m**) of Nes-cre^{ERT2};iDTA and control mice 20 weeks after transplantation of MPN BM cells ($n = 3$). **n**, Frequency of lin⁺sca-1⁺c-kit⁺ (LSK) haematopoietic progenitors in BM nucleated cells (BMNC) of Nes-cre^{ERT2};Cxcl12^{fl/fl} ($n = 5$) and control littermates ($n = 7$) 30 weeks after transplantation with MPN BM cells and 24 weeks after tamoxifen treatment. **c–e**, **j**, 6–8 weeks after pIpC treatment. Scale bars (**d**, **h**), 200 μ m. Magnification (**l**, **m**) $\times 100$. Mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$ (unpaired two-tailed *t* test). BM, bone marrow. C, control (disease-free) mice.

¹Stem Cell Niche Pathophysiology Group, Centro Nacional de Investigaciones Cardiovasculares (CNIC), 28029 Madrid, Spain. ²University Hospital Basel, CH-4031 Basel, Switzerland. ³Department of Haematology, IBSAL-Hospital Universitario de Salamanca, 37007 Salamanca, Spain. ⁴National Taiwan University, Taipei 10002, Taiwan.

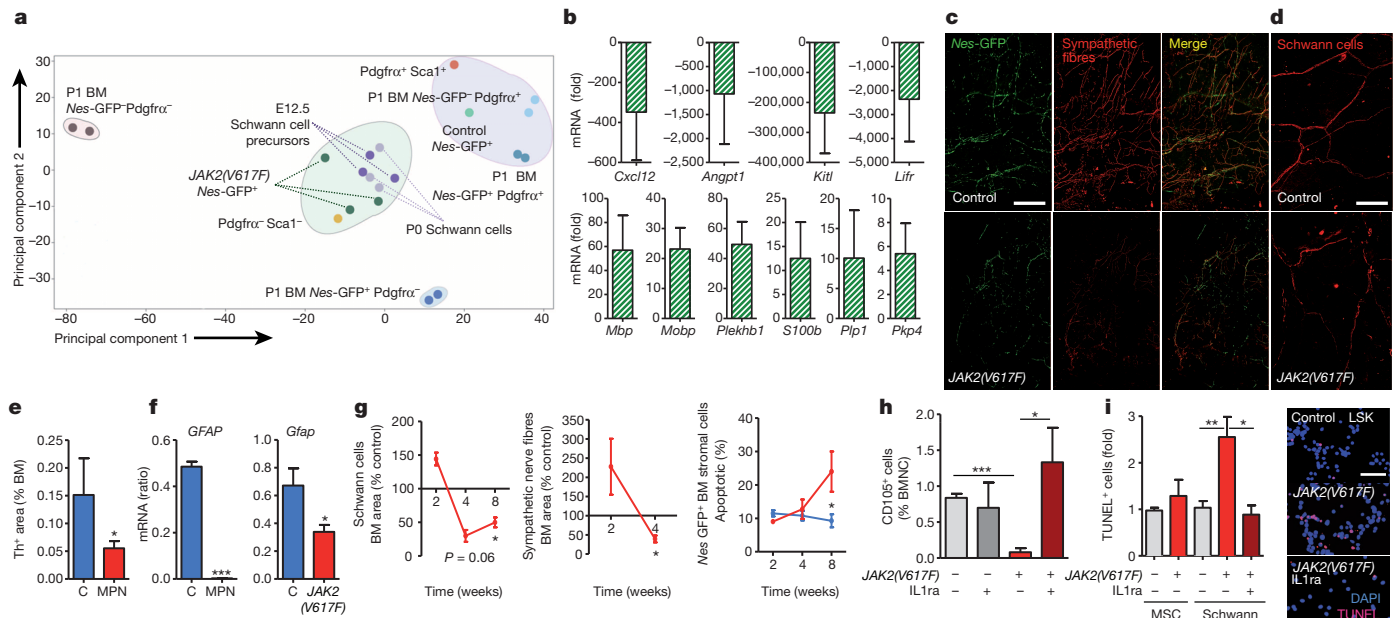
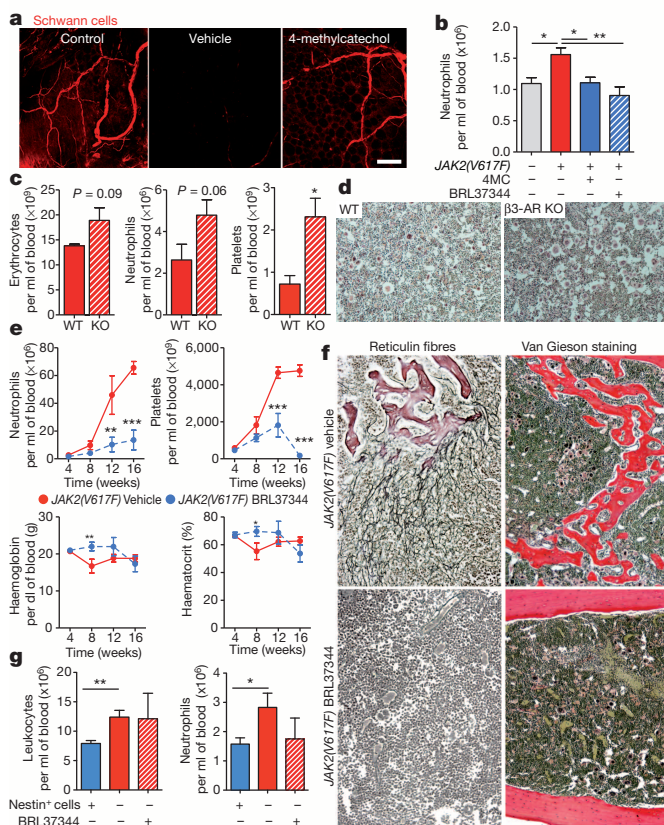


Figure 2 | BM Schwann cell death triggered by HSPC-derived interleukin-1 β precedes nestin⁺ MSC apoptosis in MPN. **a**, Principal component analysis of control and MPN BM CD45⁺CD31⁺Ter119⁺Nes-GFP⁺ cells compared with mesenchymal stromal (adult BM Pdgfrα⁺Sca1⁺ and neonatal BM Pdgfrα⁺Nes-GFP⁺ cells; purple shaded area) and Schwann cells (embryonic day 12.5 (E12.5) Schwann cell precursors and neonatal Schwann cells; green shaded area; see Methods). **b**, qPCR validation (*n* = 3). **c**, Nes-gfp skull BM 5 weeks after transplantation with control or MPN BM cells; fluorescent signals of GFP (green) and sympathetic nerve fibres detected with anti-tyrosine hydroxylase (Th) antibodies (red). **d**, Immunostaining of glial fibrillary acidic protein (Gfap) to visualize Schwann cells in control and MPN BM. Scale bar, 200 μm (c), 100 μm (d). **e**, Quantification of Th⁺ fibres in BM from controls (*n* = 2) and MPN

patients (*n* = 16). **f**, GFAP mRNA expression in BM cells of controls (*n* = 2), MPN patients (*n* = 11) and mice (*n* = 4). **g**, Time-course analyses of Schwann cells (Gfap), sympathetic fibres (Th) and Nes-GFP⁺ cell apoptosis in Nes-gfp;Mx1-cre;JAK2(V617F)⁺ and control mice 2, 4 (*n* = 5) and 8 (*n* = 7) weeks after pIpC treatment. **h**, Frequency of BM CD45⁺CD31⁺Ter119⁺CD105⁺ cells after 18-week interleukin-1 receptor antagonist (IL1ra) treatment (*n* = 5). **i**, Apoptotic rate of MSCs and Schwann cells *in vitro* derived from neonatal BM Nes-GFP⁺ cells and co-cultured for 24 h with control or MPN adult BM lin⁺sca-1⁺c-kit⁺ (LSK) cells (± 200 ng ml⁻¹ IL1ra) (*n* = 3 experiments). TUNEL staining (pink) of Schwann cells; nuclei were counterstained with DAPI (blue). Scale bar, 100 μm. Mean \pm s.e.m. **P* < 0.05; ***P* < 0.01; ****P* < 0.001 (unpaired two-tailed *t* test). HSPC, haematopoietic stem and progenitor cells.



mice that developed MPN upon pIpC treatment^{19,20} (Fig. 1c). MPN mice were intercrossed with a Nes-gfp line to label MSCs⁵ with green fluorescent protein (GFP). Both compound mutant mice and Nes-gfp animals transplanted with mutant BM cells developed MPN and had fewer BM MSCs defined by GFP, surface marker expression and functional assays (Fig. 1d–g and Extended Data Fig. 1a–f). Because MSC loss was concomitant with incipient BM fibrosis, we conducted long-term *in vivo* lineage-tracing studies to determine whether nestin⁺ MSCs differentiate into fibroblasts or osteoblasts in MPN, thereby contributing to the stromal changes in these mice^{19,20} (Extended Data Fig. 1g–j). Unexpectedly, the vascular patterns of GFP⁺ cells were similar to those in unaffected

Figure 3 | Treatment with β₃-adrenergic agonist or neuroprotective drug blocks MPN progression. **a**, **b**, Neurotrophic rescue of BM Schwann cells blocks MPN progression. Wild-type mice transplanted with MPN BM cells were treated over a month with BRL37344, the neuroprotective drug 4-methylcatechol (4MC) or vehicle. **a**, Skull BM immunostaining of glial fibrillary acidic protein (red) to visualize Schwann cells (*n* = 5); scale bar, 100 μm. **b**, Circulating neutrophils (control, JAK2(V617F) 4MC, JAK2(V617F) BRL37344, *n* = 5; JAK2(V617F) vehicle, *n* = 9). **c**, Circulating erythrocytes, neutrophils and platelets 16 weeks after transplantation of MPN BM cells into β₃-adrenergic receptor-deficient (KO, *n* = 8) and wild-type (WT) mice (*n* = 7). **d**, Van Gieson stainings of femoral BM of mice in **c**. **e**–**g**, Compensation of BM sympathetic damage by selective sympathomimetic drugs blocks MPN progression and prevents fibrosis. **e**, Blood counts of WT mice transplanted with MPN BM cells and chronically treated with selective β₃-adrenergic agonist (BRL37344) or vehicle (*n* = 5). **f**, Reticulin and Van Gieson stainings of femoral BM of mice in **e** (magnification, $\times 200$). **g**, *In vivo* depletion of nestin⁺ cells reduces the therapeutic effect of BRL37344. Blood counts of *i*DTA control (*n* = 5) and Nes-cre^{ERT2};iDTA mice treated with vehicle (*n* = 4) or BRL37344 (*n* = 5) for 6 weeks. Drug treatments in **a**, **b**, **e**, **f** started 4 weeks after transplantation; treatment in **g** started 2 weeks after transplantation combined with tamoxifen (4 weeks). Mean \pm s.e.m. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (unpaired two-tailed *t* test).

Nes-gfp mice (Fig. 1h, i). Thus, as was recently reported in BCR-ABL⁺ MPN²¹, *Nes-GFP*⁺ cells might generate excessive fibroblasts and osteoblasts. Nestin⁺ MSC reduction was instead explained by a threefold increased apoptotic rate in mutant mice (Fig. 1j), and was not prevented by the JAK inhibitor ruxolitinib (Extended Data Fig. 2).

To determine whether nestin⁺ MSC death could in turn stimulate MPN progression, we selectively depleted nestin⁺ cells *in vivo*. This depletion did not affect mature BM Schwann cells, reported to express nestin²², but reduced MSCs, associated with increased white and red blood cells (Fig. 1k and Extended Data Fig. 3a–e). BM sections revealed excessive fibroblasts and bone formation, not yet detectable in control mice (Fig. 1l), consistent with nestin⁺ cells not generating fibroblasts or bone cells in MPN. Disease acceleration following nestin⁺ cell depletion also manifested as severe spleen infiltration, still absent in control mice (Fig. 1m).

At an early disease stage, most primitive HSCs showed highest expansion, leading to increased haematopoietic progenitors in BM, peripheral blood and spleen. The chemokine *Cxcl12* regulates HSC migration and quiescence^{23,24} and is highly expressed by nestin⁺ MSCs⁵. Early HSC mobilisation correlated with decreased BM *Cxcl12*, consistent with MSC reduction. In addition, *Cxcl12* expression dropped 70-fold in MPN BM *Nes-GFP*⁺ cells (Extended Data Fig. 3f–k). Deletion of *Cxcl12* (ref. 24) in nestin⁺ cells *in vivo* increased BM haematopoietic progenitors and circulating platelets (Fig. 1n and Extended Data Fig. 3l). MSC-derived *Cxcl12* can thus negatively regulate *JAK2(V617F)*⁺ HSC expansion.

To better understand BM nestin⁺ cell alterations, genome-wide expression was profiled by next-generation sequencing. Expression of MSC and HSC-related genes was lower in MPN *Nes-GFP*⁺ cells, which instead showed enrichment in Schwann cell genes and neural-related functional categories (Extended Data Figs 4a–f and 10). Principal component analyses of independent biological samples compared with publicly available data showed that control *Nes-GFP*⁺ cells were closest to mesenchymal progenitors, whereas MPN *Nes-GFP*⁺ clustered away from them and close to Schwann cells (Fig. 2a). These changes, confirmed by quantitative PCR (Fig. 2b), suggested an altered HSC niche's neural component in MPN.

Sympathetic nerve fibres and ensheathing Schwann cells, adjacent to distinctive *Nes-GFP*⁺ cells, and *GFAP* mRNA expression were markedly reduced in BM of MPN patients and mice (Fig. 2c–f and Extended Data Fig. 4g–i). Time-course analysis showed that BM neural damage precedes *Nes-GFP*⁺ cell apoptosis (Fig. 2g), indicating that sympathetic neuropathy could sensitise nestin⁺ cells to cell death triggered by mutant cells. Multiplex ELISA detected early increased interleukin-1 β in MPN BM (Extended Data Fig. 5a). This cytokine and its activating enzyme caspase-1 were expressed by monocytes, as previously reported²⁵, but also by haematopoietic progenitors (Extended Data Fig. 5b–d). Compared with BM *Nes-GFP*⁺ stromal cells, mRNA levels of interleukin-1 receptor and its antagonist were 10- and 1,000-fold higher in *Nes-GFP*⁺ cells, respectively, and specifically upregulated in *Nes-GFP*⁺ cells in MPN (Extended Data Fig. 5e, f). Therefore we chronically treated mice with an antagonist of interleukin-1 receptor. This treatment reduced platelet counts and increased BM MSC frequency, associated with reduced caspase-1 mRNA expression in haematopoietic progenitors (Fig. 2h and Extended Data Fig. 5g–j). We studied whether *JAK2(V617F)*⁺ HSCs might directly cause BM Schwann cell death. Unlike MSCs, BM-derived Schwann cells co-cultured with *JAK2(V617F)*⁺ haematopoietic progenitors showed threefold higher apoptotic rate, which was blocked by interleukin-1 receptor antagonist (Fig. 2i). Together, these data suggest that HSC-derived interleukin-1 β contributes to neuroglial damage, which compromises MSC survival. We therefore investigated whether sympathetic neuropathy might underlie HSC niche alterations and thus represent a therapeutic target in MPN.

We treated symptomatic MPN mice with the neuroprotective agent 4-methylcatechol, which can protect BM sympathetic nerve fibres during chemotherapy²⁶. Schwann cells were preserved in 4-methylcatechol-treated mice, associated with prevented neutrophilia (Fig. 3a, b). Sympathetic nerve fibres regulate BM HSC traffic via β_3 -adrenergic receptor activation in nestin⁺ MSCs⁵⁶. This receptor is not expressed in normal or *JAK2(V617F)*⁺ haematopoietic cells (Extended Data Fig. 6a). Disease development was

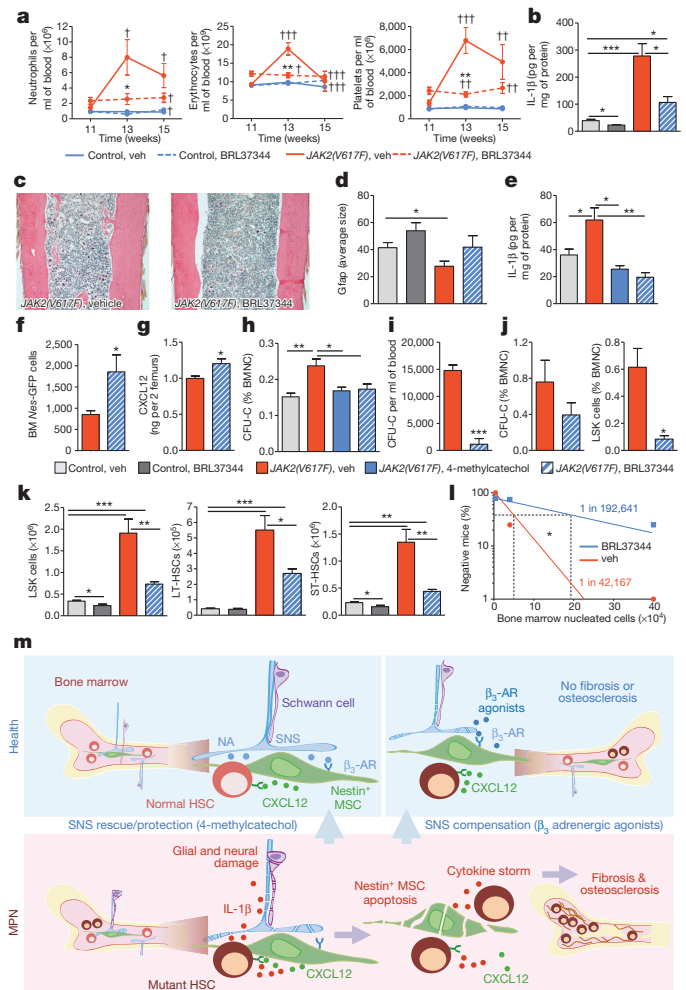


Figure 4 | Compensation of BM neuropathy rescues MSCs and prevents mutant HSC expansion in MPN. **a–d**, Efficacy of BRL37344 treatment in advanced MPN; WT mice transplanted with control or MPN BM cells were treated with BRL37344 or vehicle (veh) upon thrombocytosis ($869 \pm 23 \times 10^6$ and $1,968 \pm 264 \times 10^6$ platelets per ml of blood, respectively; $n = 5$). **a**, Blood counts; $\dagger P < 0.05$, $\dagger\dagger P < 0.01$, $\dagger\dagger\dagger P < 0.001$ vs vehicle-treated control mice. **b**, IL-1 β content in BM supernatant. **c**, Van Gieson staining of femoral BM sections (magnification, $\times 100$). **d**, Quantification of BM glial fibrillary acidic protein (Gfap) immunostaining. **e–l**, Sympathomimetic drug restores BM *Cxcl12* levels and prevents HSC expansion and mobilisation. WT mice transplanted with control or MPN BM cells were treated 4 weeks with BRL37344, 4-methylcatechol or vehicle over 4 (**e**, **h**), 8 (**f**) or 16 weeks (**g**, **i**, **j**). **e**, IL-1 β content in BM supernatant (control vehicle, $n = 10$; *JAK2(V617F)* vehicle, $n = 11$; *JAK2(V617F)* 4-methylcatechol, *JAK2(V617F)* BRL37344, $n = 5$). **f**, BM CD45⁺CD31⁺Ter119⁺*Nes-GFP*⁺ cells in vehicle ($n = 8$) and BRL-treated ($n = 10$) MPN mice. **g**, *Cxcl12* content in BM supernatant ($n = 5$). **h–j**, Frequency of colony-forming units (CFU-C) in BM nucleated cells (BMNC) (**h**, $n = 5$; **j**, $n = 3$) and blood (**i**), and BM fraction of lin[−]sca-1⁺c-kit⁺ (LSK) cells ($n = 3$) (**j**). **k**, LSK cells, long-term (LT-) and short-term (ST-) HSCs ($n = 5$) in BM from mice in **a–d**. **l**, Reduction of leukaemic stem cells in BRL37344-treated mice. CD45.2 WT mice were transplanted with BM cells from CD45.1 mice together with BM cells from MPN mice treated with vehicle or BRL37344 ($n = 5$). Frequency of mice with $<50\%$ donor LSK cell chimaerism 16 weeks after transplantation is plotted against tested cell number. MPN-initiating cell frequency is indicated. $*P < 0.05$, Pearson chi-squared t test. **m**, Model illustrating HSC niche alterations and rescue in MPN. MPN, myeloproliferative neoplasm; HSC, haematopoietic stem cell; SNS, sympathetic nervous system; MSC, mesenchymal stem cell; NA, noradrenaline; AR, adrenergic receptor; C, control (disease-free mice). **a–k**, Mean \pm s.e.m. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (unpaired two-tailed t test).

accelerated in mice lacking the β_3 -adrenergic receptor (Fig. 3c, d), uncovering a protective role of this receptor in MPN. Symptomatic mice were chronically treated with a selective β_3 -adrenergic agonist (BRL37344) to compensate for deficient sympathetic stimulation of nestin⁺ MSCs. Notably, BRL37344 treatment prevented MPN-associated neutrophilia and thrombocytosis, and delayed red blood cell reduction, but did not affect blood counts in wild-type mice (Fig. 3b, e and Extended Data Fig. 6b–d). Contrasting with the severe fibrosis in vehicle-injected mice, the BM of BRL37344-treated animals was virtually devoid of excessive bone and fibroblastic tissue (Fig. 3f). These effects were HSC niche-dependent, because neither 4-methylcatechol nor BRL37344 affected the growth of cultured haematopoietic progenitors and leukocytosis was not rescued by BRL37344 in mice with nestin⁺ cell depletion (Fig. 3g and Extended Data Fig. 7a). Similarly, several MPN markers were improved by treatment with the clinically-approved β_3 -adrenergic agonist mirabegron (Extended Data Fig. 7b), albeit to a lower extent, probably owing to its poor solubility and relatively low affinity for the murine receptor. To investigate the potential therapeutic benefit when administered at more advanced stages, thrombocytotic and control mice were treated with BRL37344. This treatment reduced neutrophilia, erythrocytosis, thrombocytosis, BM interleukin-1 β , fibrosis and osteosclerosis, rescued BM Schwann cells (Fig. 4a–d) and blocked Schwann cell program activation in BM nestin⁺ cells (Extended Data Fig. 7c). MPN progression can thus be blocked by protection or rescue of BM neuroglia and by compensation of deficient sympathetic stimulation of nestin⁺ MSCs by β_3 -adrenergic agonists.

We next asked whether MPN blockade could be mediated by preservation of MSCs and their HSC regulatory function. BRL37344 reduced IL-1 β , restored Nes-GFP⁺ cell number and increased Cxcl12 levels in BM (Fig. 4e–g). Early BRL37344 or 4-methylcatechol treatment prevented mutant haematopoietic progenitor expansion (Fig. 4h and Extended Data Fig. 8). Long-term BRL37344 treatment did not compromise normal HSCs but efficiently decreased mutant haematopoietic progenitors (Fig. 4i, j and Extended Data Fig. 9), even when administered at thrombocytosis stage (Fig. 4k). Moreover, BRL37344-treated MPN mice showed 4.5-fold reduction in leukaemic stem cells (Fig. 4l).

Our findings point to mutant HSCs as the cause of BM neuroglial damage that compromises MSC survival and function, critically contributing to MPN pathogenesis (Fig. 4m). The study shows that the niche damage triggered by the mutant HSC is essential for the development of a haematopoietic malignancy previously considered to be caused by the HSC alone. Targeting HSC niche-forming MSCs and their neural regulation may pave the way to more efficient therapeutic strategies in MPN.

METHODS SUMMARY

Mx1-cre;JAK2(V617F)^{fl}, Nes-cre^{ERT2}(REF28), RCE-loxP²⁹, iDTA³⁰, Cxcl12-floxed²⁴, Adrb3^{tm1Low}, CD45.1 and CD45.2 C57BL/6j mice (Jackson Laboratories) were housed in specific-pathogen-free facilities. Protocols were approved by the Animal Welfare Ethical Committee. *In vivo* treatments, cell extraction, culture, FACS, histological analyses, ELISA, qPCR, genomic and statistical analyses are detailed in the Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 July 2013; accepted 15 April 2014.

Published online 22 June 2014.

- James, C. *et al.* A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* **434**, 1144–1148 (2005).
- Kralovics, R. *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* **352**, 1779–1790 (2005).
- Levine, R. L. *et al.* Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* **7**, 387–397 (2005).
- Baxter, E. J. *et al.* Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* **365**, 1054–1061 (2005).
- Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).

- Méndez-Ferrer, S., Lucas, D., Battista, M. & Frenette, P. S. Haematopoietic stem cell release is regulated by circadian oscillations. *Nature* **452**, 442–447 (2008).
- Raaijmakers, M. H. *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* **464**, 852–857 (2010).
- Walkley, C. R. *et al.* A microenvironment-induced myeloproliferative syndrome caused by retinoic acid receptor gamma deficiency. *Cell* **129**, 1097–1110 (2007).
- Walkley, C. R., Shea, J. M., Sims, N. A., Purton, L. E. & Orkin, S. H. Rb regulates interactions between hematopoietic stem cells and their bone marrow microenvironment. *Cell* **129**, 1081–1095 (2007).
- Colmone, A. *et al.* Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science* **322**, 1861–1865 (2008).
- Iriuchishima, H. *et al.* Neovascular niche for human myeloma cells in immunodeficient mouse bone. *PLoS ONE* **7**, e30557 (2012).
- Pikman, Y. *et al.* *MPLW515L* is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med.* **3**, e270 (2006).
- Pardanani, A. D. *et al.* *MPL515* mutations in myeloproliferative and other myeloid disorders: a study of 1182 patients. *Blood* **108**, 3472–3476 (2006).
- Klampfl, T. *et al.* Somatic mutations of calreticulin in myeloproliferative neoplasms. *N. Engl. J. Med.* **369**, 2379–2390 (2013).
- Nangalia, J. *et al.* Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N. Engl. J. Med.* **369**, 2391–2405 (2013).
- Shih, A. H., Abdel-Wahab, O., Patel, J. P. & Levine, R. L. The role of mutations in epigenetic regulators in myeloid malignancies. *Nature Rev. Cancer* **12**, 599–612 (2012).
- Vannucchi, A. M. *et al.* Mutations and prognosis in primary myelofibrosis. *Leukemia* **27**, 1861–1869 (2013).
- Isern, J. *et al.* Self-renewing human bone marrow mesospheres promote hematopoietic stem cell expansion. *Cell Rep.* **3**, 1714–1724 (2013).
- Tiedt, R. *et al.* Ratio of mutant *JAK2-V617F* to wild-type *Jak2* determines the MPD phenotypes in transgenic mice. *Blood* **111**, 3931–3940 (2008).
- Kubovcakova, L. *et al.* Differential effects of hydroxyurea and INC424 on mutant allele burden and myeloproliferative phenotype in a *JAK2-V617F* polycythemia vera mouse model. *Blood* **121**, 1188–1199 (2013).
- Schepers, K. *et al.* Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. *Cell Stem Cell* **13**, 285–299 (2013).
- Yamazaki, S. *et al.* Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* **147**, 1146–1158 (2011).
- Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
- Tzeng, Y. S. *et al.* Loss of Cxcl12/Sdf-1 in adult mice decreases the quiescent state of hematopoietic stem/progenitor cells and alters the pattern of hematopoietic regeneration after myelosuppression. *Blood* **117**, 429–439 (2011).
- Rameshwar, P., Denny, T. N., Stein, D. & Gascon, P. Monocyte adhesion in patients with bone marrow fibrosis is required for the production of fibrogenic cytokines. Potential role for interleukin-1 and TGF-beta. *J. Immunol.* **153**, 2819–2830 (1994).
- Lucas, D. *et al.* Chemotherapy-induced bone marrow nerve injury impairs hematopoietic regeneration. *Nature Med.* **19**, 695–703 (2013).
- Mignone, J. L., Kukekov, V., Chiang, A. S., Steindler, D. & Enikolopov, G. Neural stem and progenitor cells in nestin-GFP transgenic mice. *J. Comp. Neurol.* **469**, 311–324 (2004).
- Balordi, F. & Fishell, G. Mosaic removal of hedgehog signaling in the adult SVZ reveals that the residual wild-type stem cells have a limited capacity for self-renewal. *J. Neurosci.* **27**, 14248–14259 (2007).
- Sousa, V. H., Miyoshi, G., Hjerling-Leffler, J., Karayannis, T. & Fishell, G. Characterization of Nkx6-2-derived neocortical interneuron lineages. *Cereb. Cortex* **19** (Suppl. 1), i1–i10 (2009).
- Brockschneider, D., Pechmann, Y., Sonnenberg-Riethmacher, E. & Riethmacher, D. An improved mouse line for Cre-induced cell ablation due to diphtheria toxin A, expressed from the Rosa26 locus. *Genesis* **44**, 322–327 (2006).

Acknowledgements We thank S. Martín-Salamanca, A. M. Martín, A. B. Ricote, J. M. Ligos, S. Bartlett and the CNIC Comparative Medicine, Genomics and Bioinformatics Units for assistance; members of S.M.-F. and B. Ibáñez groups and M. García-Fernández for data discussion; G. Enikolopov, G. Fishell and D. Riethmacher for providing mice. This work was supported by Fundación CNIC, Spanish Ministry of Economy and Competitiveness (TerCel, Spanish Cell Therapy Network; Plan Nacional SAF-2011-30308) and ConSEPOC-Comunidad de Madrid S2010/BMD-2542 grants to S.M.-F.; Ramón y Cajal Program grants RYC-2009-04703/2011-09726 and Marie Curie grants FP7-PEOPLE-2011-RG-294262/294096 to A.S.-A. and S.M.-F.; Swiss National Science Foundation 310000-120724/1, 32003BB, 135712/1 and Swiss Cancer League KLS-02398-02-2009 grants to R.C.S.; A.S.-A. received a Research Fellowship from the European Hematology Association. S.M.-F. is supported in part by an International Early Career Scientist grant of the Howard Hughes Medical Institute.

Author Contributions L.A. designed and performed most experiments and analyses, prepared figures and wrote the manuscript. A.S.-A. performed *in vivo* transplantation assays. D.M.-P. performed qPCR, genome-wide expression analyses and together with X.L. provided technical assistance. J.I. performed neonatal cell isolation and culture. X.L. and A.T. performed histological analyses. S.M. provided human samples. Y.-S.T. and D.-M.L. provided Cxcl12-floxed mice. P.L., J.S. and R.C.S. provided samples, designed experiments and human studies. S.M.-F. designed the overall study, supervised the experiments and wrote the manuscript.

Author Information The gene expression data have been deposited in the Gene Expression Omnibus (GEO) databank (<http://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE55802. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M.-F. (smendez@cnic.es).

PVT1 dependence in cancer with MYC copy-number increase

Yuen-Yi Tseng¹, Branden S. Moriarity^{2*}, Wuming Gong^{1,2*}, Ryutaro Akiyama^{1,3}, Ashutosh Tiwari^{2†}, Hiroko Kawakami^{1,3}, Peter Ronning¹, Brian Reuland¹, Kacey Guenther¹, Thomas C. Beadnell⁴, Jaclyn Essig¹, George M. Otto², M. Gerard O'Sullivan², David A. Largaespada^{1,2}, Kathryn L. Schwertfeger^{2,4*}, York Marahrens^{1*}, Yasuhiko Kawakami^{1,3*} & Anindya Bagchi^{1,2}

'Gain' of supernumerary copies of the 8q24.21 chromosomal region has been shown to be common in many human cancers^{1–13} and is associated with poor prognosis^{7,10,14}. The well-characterized myelocytomatosis (MYC) oncogene resides in the 8q24.21 region and is consistently co-gained with an adjacent 'gene desert' of approximately 2 megabases that contains the long non-coding RNA gene *PVT1*, the *CCDC26* gene candidate and the *GSDMC* gene. Whether low copy-number gain of one or more of these genes drives neoplasia is not known. Here we use chromosome engineering in mice to show that a single extra copy of either the *Myc* gene or the region encompassing *Pvt1*, *Ccdc26* and *Gsdmc* fails to advance cancer measurably, whereas a single supernumerary segment encompassing all four genes successfully promotes cancer. Gain of *PVT1* long non-coding RNA expression was required for high MYC protein levels in 8q24-amplified human cancer cells. *PVT1* RNA and MYC protein expression correlated in primary human tumours, and copy number of *PVT1* was co-increased in more than 98% of MYC-copy-increase cancers. Ablation of *PVT1* from MYC-driven colon cancer line HCT116 diminished its tumorigenic potency. As MYC protein has been refractory to small-molecule inhibition, the dependence of high MYC protein levels on *PVT1* long non-coding RNA provides a much needed therapeutic target.

To determine whether low copy-number gain of *MYC* and/or other genetic element(s) in the 8q24.21 region (*PVT1*, *CCDC26* and *GSDMC*; Fig. 1a and Extended Data Fig. 1) promotes cancer, chromosome engineering¹⁵ (Extended Data Fig. 2) was used to derive mice bearing single-copy gain of (1) *Myc* alone (Fig. 1b and Extended Data Fig. 2a), (2) *Pvt1*, *Ccdc26* and *Gsdmc* (Fig. 1c and Extended Data Fig. 2b) and (3) the entire 2-Mb *Myc*, *Pvt1*, *Ccdc26*, *Gsdmc* syntenic region (Fig. 1d and Extended Data Fig. 2c). All three strains were viable and fertile and showed no overt developmental defect. In human breast and ovarian cancers, gain of 8q24 is often accompanied by *ERBB2* amplification^{16,17}. Accordingly, each mutant line was bred to *MMTVneu* transgenic mice¹⁸ and mammary tumour latency was examined. *gain(Myc)*, *MMTVneu* mice developed mammary tumours at essentially the same median latency (345 days) as *MMTVneu* mice (Fig. 1e), indicating that a single supernumerary *Myc* gene is insufficient to promote *MMTVneu*-driven cancer. Tumour latency of *gain(Pvt1, Ccdc26, Gsdmc)*, *MMTVneu* mice was also indistinguishable from *MMTVneu* mice (316 days; Fig. 1e). In contrast, *gain(Myc, Pvt1, Ccdc26, Gsdmc)*, *MMTVneu* mice showed shorter mammary tumour latency (224 days; Fig. 1e) and increased penetrance (90%) compared with the other three genotypes (40–60%). Compared with age-matched *MMTVneu* adenomas, *gain(Myc, Pvt1, Ccdc26, Gsdmc)*, *MMTVneu* tumours presented a high mitotic index involving locally invasive solid masses that invaded adjacent thin-walled blood vessels, consistent with adenocarcinomas (Extended Data Fig. 3a–c). Three out of ten *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mice, but none of *gain(Myc)*, *gain*

(*Pvt1, Ccdc26, Gsdmc*) and wild-type mice ($n = 10$ for each genotype), developed metastasis in the absence of *MMTVneu* in a limited-lifespan study (500 days), indicating that *gain(Myc, Pvt1, Ccdc26, Gsdmc)* contributes to spontaneous metastasis in older mice, albeit with low penetrance (Extended Data Fig. 3d, e). These data demonstrate that gain of a single copy of *Myc* cooperates with one or more genetic element(s) in the *Pvt1/Ccdc26/Gsdmc* region to promote mammary tumorigenesis.

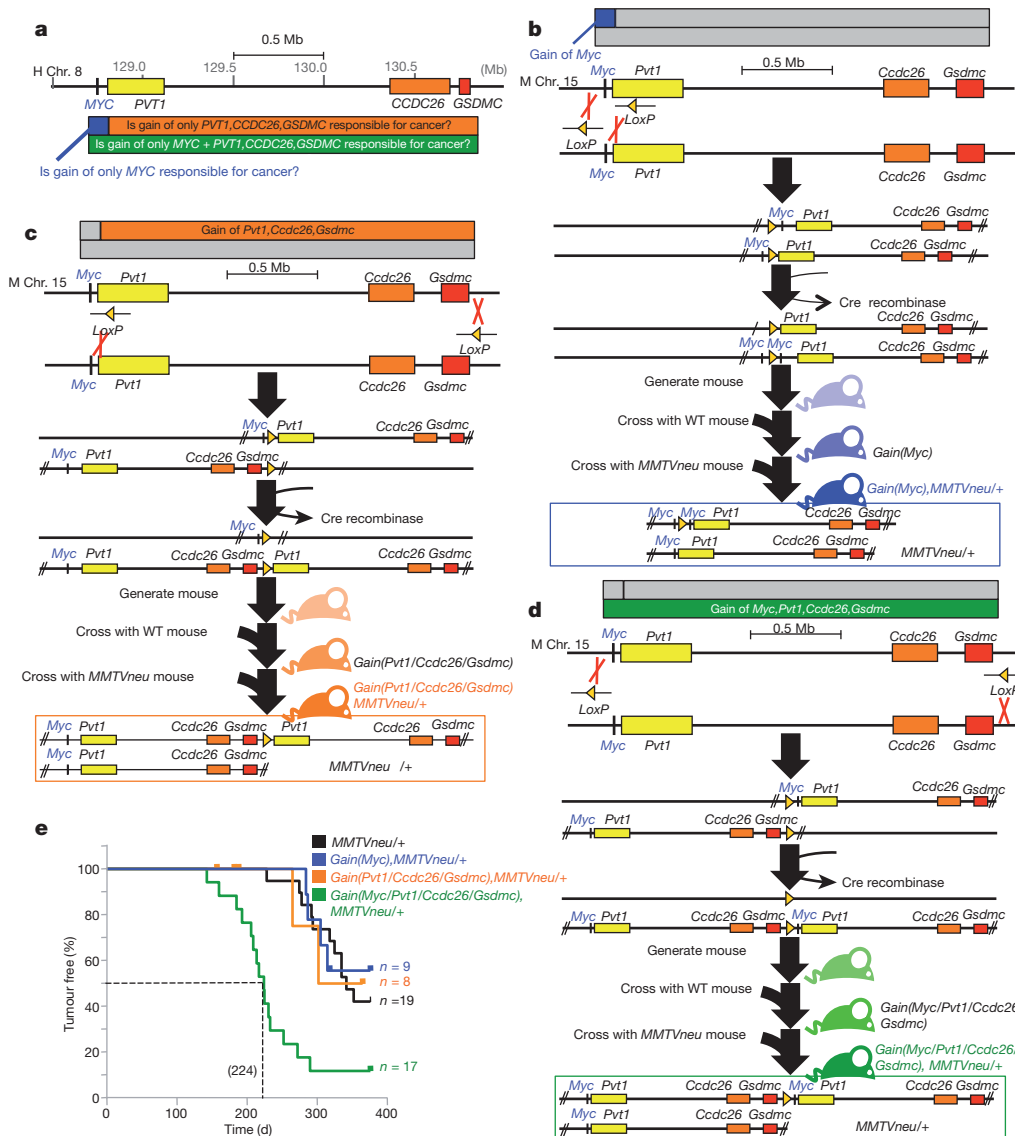
The mammary epithelium of 10-week-old *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mice showed pre-cancerous properties in the absence of *MMTVneu*, including elevated levels of γ -H2AX (a marker of DNA damage; Fig. 2a), p53 (a key mediator of cellular stress response) and phospho-ERK1/2 (a pro-survival signalling molecule; Extended Data Fig. 4a, b). Accordingly, *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mammary epithelium showed increased DNA replication (Fig. 2b), reduced oestrogen receptor- α expression¹⁹ (Extended Data Fig. 4c) and 5.5 times more lateral ductal branching than wild type (Fig. 2c). *gain(Myc)* glands showed a more modest two-fold increase in branching compared with wild type. Furthermore, *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mice showed aberrant differentiation, including a precocious alveolar-like phenotype and co-expression of keratin-14 (a myoepithelial marker) with keratin-8, in luminal epithelial cells (Extended Data Fig. 4d, e). Crossing *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mice with mice carrying a deletion of precisely the same region resulted in mice with two copies of this region on one chromosome 15 homologue and no copy on the other, which completely abolished aberrant cellular proliferation (Fig. 2d) and excessive branching (Fig. 2e). Three copies of the *Myc/Pvt1/Ccdc26/Gsdmc* region, therefore, produced pre-cancerous phenotypes.

We next sought to identify genetic elements driving *gain(Myc, Pvt1, Ccdc26, Gsdmc)* neoplasia. *Ccdc26* transcript has not been detected in mouse, and *Gsdmc* expression was low in mouse mammary tissue with no differences in *Gsdmc2-4* transcript levels across genotypes (Extended Data Fig. 5). Consequently, we focused on *Myc* and *Pvt1*. Cells from two independent *gain(Myc, Pvt1, Ccdc26, Gsdmc)*, *MMTVneu* mammary tumours were transfected with short interfering RNAs (siRNAs) to knockdown *Myc* (siMyc), *Pvt1* (siPvt1) or both, and grown in three-dimensional culture. siMyc had no effect on *Pvt1* RNA levels, and siPvt1 had no effect on *Myc* messenger RNA (mRNA) (Fig. 3a). Depletion of either *Myc* or *Pvt1* RNA resulted in approximately 60% reduction in proliferation as measured by Ki67 staining (Fig. 3b). Surprisingly, simultaneous knockdown of both *Myc* and *Pvt1* failed to reduce proliferation below individual knockdown of either gene (Fig. 3b), suggesting that *Myc* and *Pvt1* promote proliferation through the same pathway.

Analysis of *Myc* and *Pvt1* expression showed that both *Gain(Myc)* and *gain(Myc, Pvt1, Ccdc26, Gsdmc)* mammary glands showed approximately 3.5-fold more *Myc* mRNA than wild type or *gain(Pvt1, Ccdc26, Gsdmc)* (Fig. 3c). *Pvt1* mRNA levels were approximately 1.5 times higher in *gain(Pvt1, Ccdc26, Gsdmc)* than wild type, as expected, but were surprisingly

¹Department of Genetics, Cell Biology and Development, University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455, USA. ²Masonic Cancer Center, University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455, USA. ³Stem Cell Institute, University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455, USA. ⁴Department of Laboratory Medicine and Pathology, University of Minnesota, Twin Cities, Minneapolis, Minnesota 55455, USA. [†]Present address: Center for Bio-Design, Translational Health Science and Technology Institute, Gurgaon 122016, India.

*These authors contributed equally to this work.



tenfold elevated in *gain(Myc, Pvt1, Cdc26, Gsdmc)* and fourfold increased in *gain(Myc)* mammary glands (Fig. 3c). Because MYC protein is a transcriptional activator of *PVT1* (refs 20, 21), this result suggested higher levels of Myc protein and/or activity in *gain(Myc, Pvt1, Cdc26, Gsdmc)*. Threefold higher Myc protein levels were observed in *gain(Myc, Pvt1, Cdc26, Gsdmc)* mammary gland compared with the other three strains (Fig. 3d). Thus, a third copy of *Myc + Pvt1* increased *Pvt1* transcription and Myc protein in mice.

To extend these findings to high-8q24.21 copy cancers, we examined human breast-cancer cell lines SK-BR-3 and MDA-MB-231, which harbour high copy gains³. siMYC and siPVT1 had no effect on PVT1 and MYC RNA, respectively (Fig. 3e), but reduced proliferation to similar extents with combined siMYC + siPVT1 failing to reduce proliferation further (Extended Data Fig. 6a, b). Furthermore, siPVT1 led to suppression of MYC protein levels (Fig. 3f), verifying that MYC protein levels are dependent upon PVT1 mRNA in high-copy 8q24-gain cancer cells. Knockdown of individual microRNAs (miRNAs) encoded in the human *PVT1* introns had no detectable effect on proliferation (Extended Data Fig. 6c, d).

The half-life of MYC protein is increased in SK-BR-3 and MDA-MB-231 cells compared with non-transformed breast epithelial cells²². To determine whether MYC protein stability is PVT1-RNA dependent, SK-BR-3 and MDA-MB-231 cells were transfected with siPVT1 or control siRNA and exposed to cycloheximide. PVT1 depletion resulted in

more rapid loss of MYC protein than control cells (Fig. 3g and Extended Data Fig. 7). MYC protein degradation is promoted by phosphorylation at threonine 58 (ref. 23), and mice expressing *Myc*^{T58A} show enhanced mammary gland density, and mammary carcinoma²⁴. Although siPVT1-mediated reduction of MYC protein levels was not accompanied by changes in levels of FBW7 and AXIN1, which are key downstream components of MYC degradation, T58 phosphorylation was increased fivefold (Fig. 3h). Thus, MYC is protected from phosphorylation at the T58 residue and subsequent degradation in a PVT1-RNA-dependent manner, raising the possibility of a MYC/PVT1 ribonucleoprotein complex. Simultaneous RNA fluorescence *in situ* hybridization (FISH) to detect PVT1 and immunofluorescence for MYC showed nuclear co-localization in 85.1% of SK-BR-3 cells (Fig. 3i and Extended Data Fig. 8). Furthermore, immunoprecipitation using anti-MYC followed by PCR with reverse transcription (RT-PCR) identified PVT1 RNA in the immunoprecipitates (Fig. 3j), suggesting that PVT1 and MYC physically interact directly or indirectly. Whether specific PVT1 isoform(s), CCDC26 and GSDMC promote malignancy remains to be investigated. Together, these results indicate that gain of the *Myc/MYC* gene alone fails to increase MYC protein levels but that co-gain of the *Pvt1/PVT1* gene disrupts Myc/MYC instability, resulting in increased protein levels and enhanced proliferation.

If *PVT1* copy increase is critical for high MYC protein levels, then co-gain of *PVT1* should be mandatory in MYC-driven cancer. Analysis

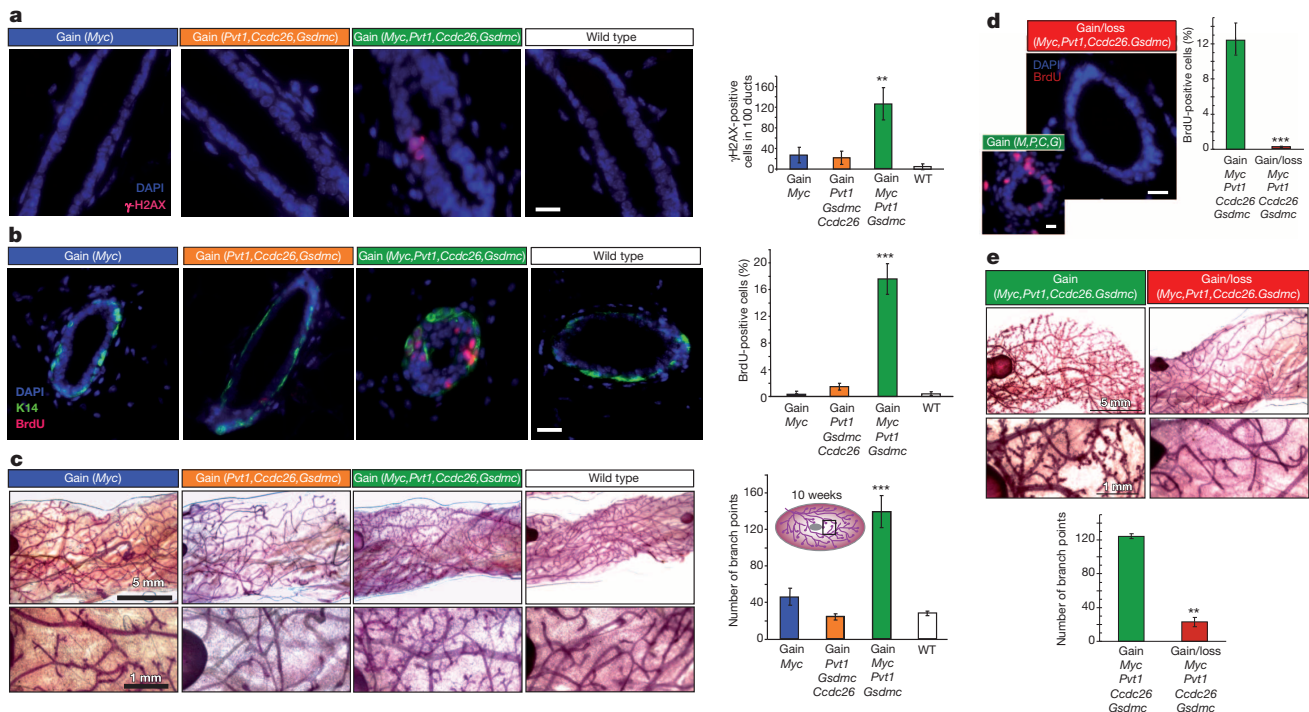


Figure 2 | Pre-cancerous phenotypes in mouse *gain(Myc,Pvt1,Ccdc26,Gsdmc)* mammary glands. **a, b,** Fluorescence images and quantification of γ -H2AX foci (**a**) and BrdU-incorporation (**b**) in mammary ducts of indicated genotype ($n = 3$). **c,** Wholemount analysis of mammary glands (higher magnification, bottom). Inset, schematic of mammary gland. Branch points were enumerated at a 25-mm² area near the lymph node ($n = 3$). **d, e,** Rescue of aberrant

proliferation (**d**), and enhanced lateral branching (**e**) in the *gain(Myc,Pvt1,Ccdc26,Gsdmc)* mammary ducts by corresponding loss allele. *M, Myc; P, Pvt1; C, Ccdc26; G, Gsdmc* ($n = 3$). Results are shown as mean \pm s.e.m. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-tailed Student's *t*-test). Scale bar on **a, b, d**, 10 μ m; **c, e**, 1 mm, 5 mm; error bars, s.e.m.

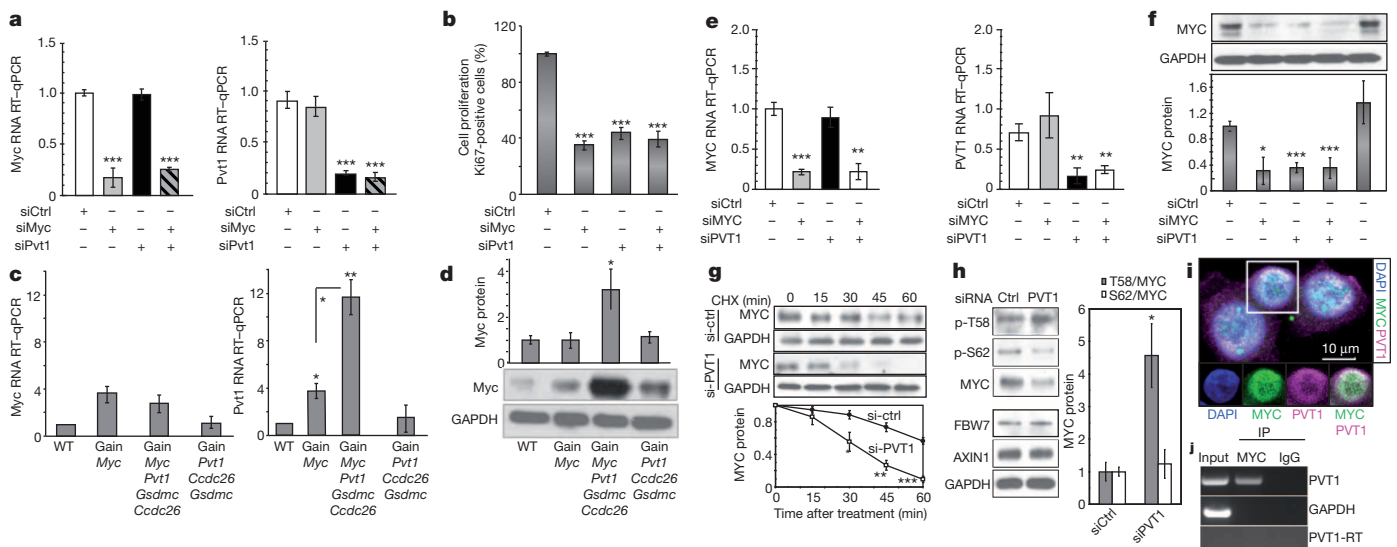


Figure 3 | *Pvt1/PVT1* co-gained with *Myc/MYC* elevates *Myc/MYC* protein levels. **a,** RT-qPCR measurement of *Myc* (left) and *Pvt1* (right) RNA levels in *gain(Myc,Pvt1,Ccdc26,Gsdmc),MMTVneu/+* mammary tumour cells transfected with indicated siRNAs ($n = 3$). **b,** Proportions of primary tumour cells positive for Ki-67 after the indicated siRNA treatments ($n = 3$). **c,** RT-qPCR of *Myc* (left) and *Pvt1* (right) transcript levels ($n = 3$), and **d,** western blot analysis (top) and quantification (bottom) of Myc protein in mammary tissue ($n = 3$). GAPDH, Glyceraldehyde 3-phosphate dehydrogenase. **e–h,** Analyses of human breast cancer cell line SK-BR-3. **e,** RT-qPCR measurement of *MYC* (left) and *PVT1* (right) transcripts in cells 48 h after transfection with the indicated siRNA(s) ($n = 3$). **f, g,** Western blot analysis of the MYC protein in SK-BR-3 after siRNA transfection ($n = 3$) (**f**), and siRNA transfection and cycloheximide (CHX)

treatment for times indicated ($n = 3$) (**g**). **h,** Western blot analysis of MYC(p-T58), MYC(p-S62), MYC, FBW7 and AXIN1 protein levels in SK-BR-3 treated with siRNAs (left). Ratios of T58/total MYC and p-S62/total MYC (right) ($n = 3$). **i,** Immunofluorescence staining of MYC (green) and RNA FISH of *PVT1* (magenta) showing nuclear co-localization of MYC and *PVT1* (white). 4',6-Diamidino-2-phenylindole (DAPI) is shown in blue. The marked cell in the upper panel is shown in the lower panels in single channels and MYC + *PVT1* overlay. **j,** RT-PCR using *PVT1* and GAPDH specific primers of total SK-BR-3 RNA (input), immunoprecipitated using MYC antibody (IP MYC) and IgG (IP IgG). *PVT1*-RT indicates samples not treated with reverse transcriptase. Results are shown as mean \pm s.e.m. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-tailed Student's *t*-test). Scale bar, 10 μ m; error bars, s.e.m.

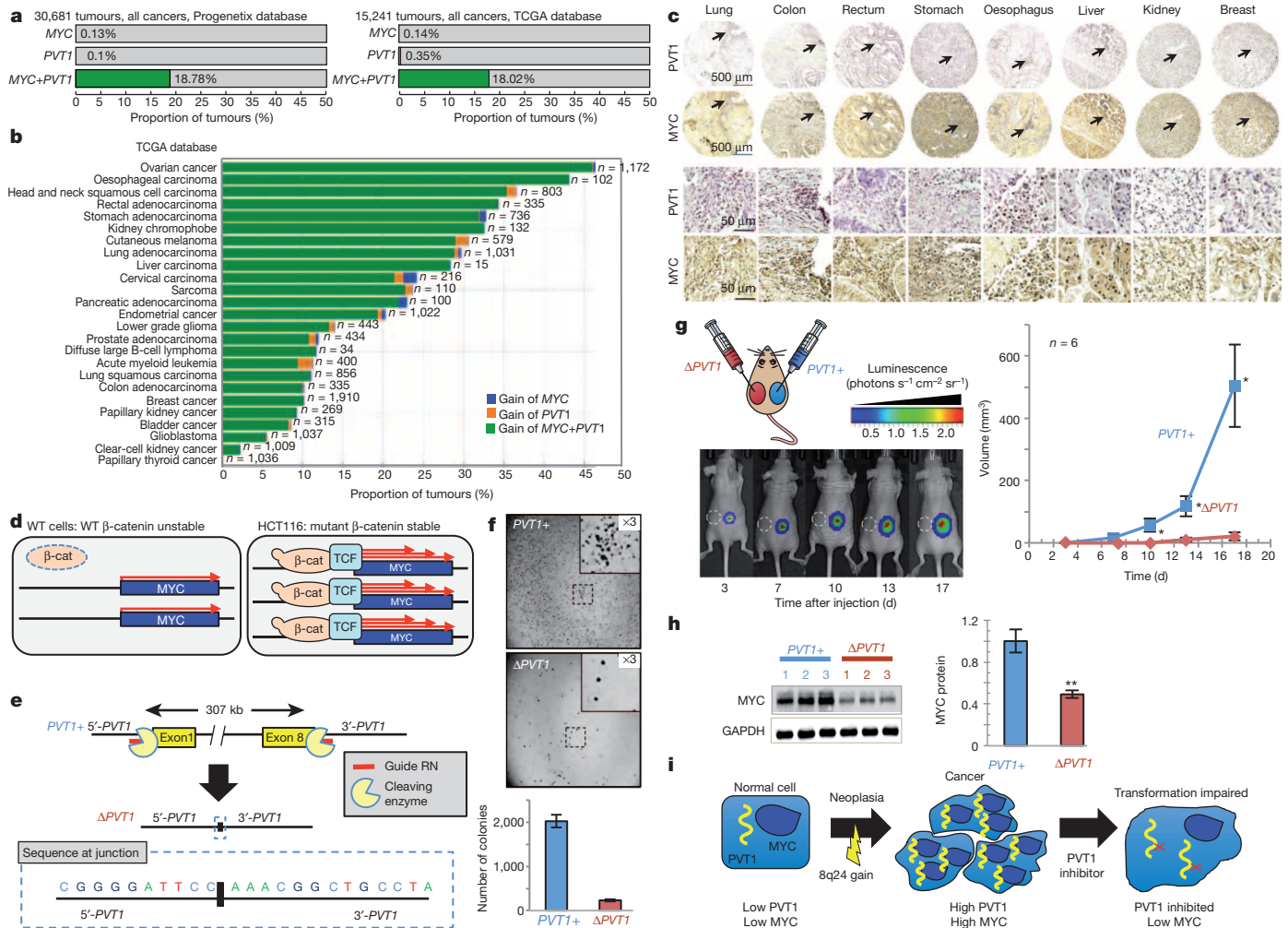


Figure 4 | PVT1 dependence in MYC-driven tumours. **a**, **b**, Proportion of all tumours harbouring gain of *MYC* but not *PVT1* (blue), *PVT1* but not *MYC* (orange) and *MYC* + *PVT1* (green) in the Progenetix (left) and TCGA databases (right) (**a**) and among different cancer types in the TCGA database (**b**). **c**, Tissue microarray analysis showing nuclear expression of PVT1 (dark purple) and MYC (dark brown) in primary human tumours. Lower panels represent $\times 10$ magnification of regions shown by arrow in the upper panel. **d**, Cartoons showing that stabilized mutant β -catenin (β -cat) upregulates *MYC* transcription through the recruitment of T-cell factor (TCF) in human colorectal cancer line HCT116. **e**, Schematic of CRISPR-mediated excision of *PVT1* to obtain the Δ PVT1 allele. DNA sequence of a PCR amplicon containing the junction sequence of the deletion product is shown. **f**, Images of colonies formed by PVT1+ and Δ PVT1 HCT116 cells in soft agar assay (top).

of 30,681 tumours from the Progenetix copy-number database showed that 18.8% (5,836 tumours) showed increased 8q24 copy number and 5,763 out of 5,836 of these tumours (98.7%) had increased copy number of both *MYC* and *PVT1* genes. Similarly, analysis of 15,241 tumours from The Cancer Genome Atlas (TCGA) database showed that 18.02% (2,821 tumours) showed 8q24 copy-number increase and that 2,746 out of 2,821 tumours (97.34%) showed co-gain of both *MYC* and *PVT1* whereas fewer than 0.15% of tumours showed increased copy number of *MYC* but not *PVT1* (Fig. 4a and Extended Data Fig. 9a). Sorting TCGA tumours by cancer type showed differences in the incidence of 8q24 copy-number increase in individual cancer types ranging from *MYC* + *PVT1* increase in almost half of ovarian and oesophageal carcinomas to essentially no *MYC* and/or *PVT1* increase among papillary thyroid tumours (Fig. 4b). Gain of *MYC* and *PVT1* was observed in 62% of 483 HER2+ breast cancer samples (Extended Data Fig. 9b). Unlike mouse, GSDMC orthologues are expressed in human mammary

tissue (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=GSDMC>). Although gain of *MYC* + *PVT1* + *CCDC26* + *GSDMC* prevailed in low copy-number gain of 8q24 segments, co-gain of *MYC* + *PVT1* (but not *CCDC26* or *GSDMC*) prevailed in high copy amplifications (Extended Data Fig. 9c, d). Co-gain of *MYC* and *PVT1*, therefore, dominated over gain of *MYC* alone across all cancer types showing 8q24 copy-number increases.

To verify these findings, we examined tissue microarrays for PVT1 RNA using *in situ* hybridization and MYC protein using immunohistochemistry on serial sections of human lung, colon, rectum, stomach, oesophagus, liver, kidney and breast tumours. Concurrent PVT1 RNA and MYC protein expressions were found in all eight cancers (Fig. 4c and Extended Data Fig. 9e, f), confirming that PVT1 RNA and the MYC protein are correlated in primary tumours.

Finally, effects of *PVT1* loss on MYC-driven tumours were assessed using HCT116 human colorectal carcinoma cells, which harbour low

copy-number 8q24 gain and stable mutant β -catenin leading to MYC overexpression (Fig. 4d)²⁵. CRISPR-associated nuclease (Cas9) was used to excise precisely all copies of the PVT1 gene (Fig. 4e and Extended Data Fig. 10a, b) generating PVT1-null (Δ PVT1) HCT116 cell lines (Extended Data Fig. 10c). Δ PVT1 lines demonstrated reduced proliferation (Extended Data Fig. 10d) and impaired colony formation in soft agar compared with PVT1+ HCT116 cells (232.2 ± 23.8 versus $2,022.3 \pm 140.7$ colonies; Fig. 4f). In xenograft studies, Δ PVT1 HCT116 cells either failed to form tumours (three out of six xenografts; Fig. 4g) or showed markedly reduced volume (three out of six xenografts; Extended Data Fig. 10e) compared with PVT1+ counterparts (Fig. 4g). Finally, MYC protein was significantly reduced ($49.1\% \pm 3.4$) in Δ PVT1 HCT116 clones compared with PVT1+ HCT116 cells (Fig. 4h). PVT1 therefore regulates MYC protein level, and bestows tumorigenic potential to an MYC-driven non-breast cancer line.

Targeting MYC directly with therapeutic interventions has proved challenging^{26–28}, thus regulation of high MYC protein levels by the PVT1 long non-coding (lncRNA) bears considerable implications for therapeutic treatment of MYC-driven cancers. Because MYC is an important transcription factor and an essential protein²⁷, direct inhibition may have severe effects on patients. Our findings indicate that PVT1 lncRNA increases MYC protein levels in 8q24-gain cancers and that loss of PVT1 RNA reduces MYC protein to more normal levels. PVT1, therefore, may be a more accessible and less deleterious target than MYC itself for curtailing MYC-driven cancers (Fig. 4i). Future studies on the role of PVT1 on the MYC protein level in cancers without supernumerary 8q24 and illuminating the molecular details of MYC/PVT1 cooperation may lead to rational drug discovery specifically targeting the MYC/PVT1 axis in human cancers.

METHODS SUMMARY

Chromosome engineering on mouse AB2.2 embryonic stem cells was as previously described²⁹ (Extended Data Fig. 2). Mouse mammary glands were analysed as described³⁰. Tissue microarray slides of human multiple-organ tumours (BC00119) were obtained from US Biomax. For copy-number analysis of TCGA tumours, data were derived from the Affymetrix Genome Wide Human SNP Array 6.0 platform from the open-access directory of TCGA (https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/). CRISPR-mediated Δ PVT1 HCT116 cells were generated using piggyBac co-transposition enrichment.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 July 2013; accepted 8 April 2014.

Published online 22 June; corrected online 6 August 2014 (see full-text HTML version for details).

- Huppi, K., Pitt, J. J., Wahlberg, B. M. & Caplen, N. J. The 8q24 gene desert: an oasis of non-coding transcriptional activity. *Front. Genet.* **3**, 69 (2012).
- Haverty, P. M., Hon, L. S., Kaminker, J. S., Chant, J. & Zhang, Z. High-resolution analysis of copy number alterations and associated expression changes in ovarian tumors. *BMC Med. Genomics* **2**, 21 (2009).
- Guan, Y. *et al.* Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clin. Cancer Res.* **13**, 5745–5755 (2007).
- van Duin, M. *et al.* High-resolution array comparative genomic hybridization of chromosome 8q: evaluation of putative progression markers for gastroesophageal junction adenocarcinomas. *Cytogenet. Genome Res.* **118**, 130–137 (2007).
- Borg, A., Baldetorp, B., Ferno, M., Olsson, H. & Sigurdsson, H. c-myc amplification is an independent prognostic factor in postmenopausal breast cancer. *Int. J. Cancer* **51**, 687–691 (1992).
- Kim, Y. H. *et al.* Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of MYC family gene amplification. *Oncogene* **25**, 130–138 (2006).
- Sato, K. *et al.* Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *J. Natl Cancer Inst.* **91**, 1574–1580 (1999).

- Lapointe, J. *et al.* Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res.* **67**, 8504–8510 (2007).
- Douglas, E. J. *et al.* Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res.* **64**, 4817–4825 (2004).
- Zitterbart, K. *et al.* Low-level copy number changes of MYC genes have a prognostic impact in medulloblastoma. *J. Neurooncol.* **102**, 25–33 (2011).
- Yamada, T. *et al.* Frequent chromosome 8q gains in human small cell lung carcinoma detected by arbitrarily primed-PCR genomic fingerprinting. *Cancer Genet. Cytogenet.* **120**, 11–17 (2000).
- Le Beau, M. M., Bitts, S., Davis, E. M. & Kogan, S. C. Recurring chromosomal abnormalities in leukemia in PML-RARA transgenic mice parallel human acute promyelocytic leukemia. *Blood* **99**, 2985–2991 (2002).
- Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529–541 (2006).
- Jain, A. N. *et al.* Quantitative analysis of chromosomal CGH in human breast tumors associates copy number abnormalities with p53 status and patient survival. *Proc. Natl Acad. Sci. USA* **98**, 7952–7957 (2001).
- Ramirez-Solis, R., Liu, P. & Bradley, A. Chromosome engineering in mice. *Nature* **378**, 720–724 (1995).
- Al-Kuray, K. *et al.* Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res.* **64**, 8534–8540 (2004).
- Park, K., Kwak, K., Kim, J., Lim, S. & Han, S. c-myc amplification is associated with HER2 amplification and closely linked with cell proliferation in tissue microarray of nonselected breast cancers. *Hum. Pathol.* **36**, 634–639 (2005).
- Guy, C. T. *et al.* Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl Acad. Sci. USA* **89**, 10578–10582 (1992).
- Saji, S. *et al.* Estrogen receptors α and β in the rodent mammary gland. *Proc. Natl Acad. Sci. USA* **97**, 337–342 (2000).
- Carramusa, L. *et al.* The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. *J. Cell. Physiol.* **213**, 511–518 (2007).
- Lin, M. *et al.* RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS ONE* **6**, e23356 (2011).
- Zhang, X. *et al.* Mechanistic insight into Myc stabilization in breast cancer involving aberrant Axin1 expression. *Proc. Natl Acad. Sci. USA* **109**, 2790–2795 (2012).
- Yeh, E. *et al.* A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nature Cell Biol.* **6**, 308–318 (2004).
- Wang, X. *et al.* Phosphorylation regulates c-Myc's oncogenic activity in the mammary gland. *Cancer Res.* **71**, 925–936 (2011).
- Morin, P. J. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutations in β -catenin or APC. *Science* **275**, 1787–1790 (1997).
- Delmore, J. E. *et al.* BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
- Darnell, J. E., Jr. Transcription factors as targets for cancer therapy. *Nature Rev. Cancer* **2**, 740–749 (2002).
- Nair, S. K. & Burley, S. K. X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112**, 193–205 (2003).
- Bagchi, A. *et al.* CHD5 is a tumor suppressor at human 1p36. *Cell* **128**, 459–475 (2007).
- Schwertfeger, K. L. *et al.* A critical role for the inflammatory response in a mouse model of preneoplastic progression. *Cancer Res.* **66**, 5676–5685 (2006).

Acknowledgements We thank A. T. Vogel for writing statistical analysis scripts; Research Animal Resources, University of Minnesota, for maintaining the mouse colony; S. Horn and L. Oseth for embryonic stem cell blastocyst injection and FISH analysis respectively. This work was supported by Masonic Cancer Center Laboratory start-up funds (to A.B.), and by grants from the Masonic Scholar Award (to A.B.), the Karen Wyckoff Rein in Sarcoma Fund (to A.B.), Translational Workgroup Pilot Project Awards by the Institute of Prostate and Urologic Cancer, University of Minnesota (to A.B.) and an American Cancer Society Institutional Research Grant (award 118198-IRG-58-001-52-IRG92, to A.B.). A.T. was supported by an Indo-US fellowship from the Indo-US Science and Technology Forum.

Author Contributions Y.Y.T. and A.B. conceptualized the research programme and designed the experiments; Y.Y.T., B.S.M., H.K., A.T., R.A., P.R., B.R., K.G., T.C.B., J.E., Y.K. and A.B. performed the experiments. Y.Y.T. and W.G. analysed the data; M.G.O. and Y.Y.T. performed the histological analyses; K.L.S., D.A.L., Y.M., Y.K. and A.B. supervised experiments and data analysis; A.B. and Y.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.B. (bagch005@umn.edu).

Putative *cis*-regulatory drivers in colorectal cancer

Halit Ongen^{1,2,3}, Claus L. Andersen⁴, Jesper B. Bramsen⁴, Bodil Oster⁴, Mads H. Rasmussen⁴, Pedro G. Ferreira^{1,2,3}, Juan Sandoval⁵, Enrique Vidal⁵, Nicola Whiffin⁶, Alexandra Planchon^{1,2,3}, Ismael Padioleau^{1,2,3}, Deborah Bielser^{1,2,3}, Luciana Romano^{1,2,3}, Ian Tomlinson⁷, Richard S. Houlston⁶, Manel Esteller^{5,8,9}, Torben F. Orntoft⁴ & Emmanouil T. Dermizakis^{1,2,3}

The *cis*-regulatory effects responsible for cancer development have not been as extensively studied as the perturbations of the protein coding genome in tumorigenesis^{1,2}. To better characterize colorectal cancer (CRC) development we conducted an RNA-sequencing experiment of 103 matched tumour and normal colon mucosa samples from Danish CRC patients, 90 of which were germline-genotyped. By investigating allele-specific expression (ASE) we show that the germline genotypes remain important determinants of allelic gene expression in tumours. Using the changes in ASE in matched pairs of samples we discover 71 genes with excess of somatic *cis*-regulatory effects in CRC, suggesting a cancer driver role. We correlate genotypes and gene expression to identify expression quantitative trait loci (eQTLs) and find 1,693 and 948 eQTLs in normal samples and tumours, respectively. We estimate that 36% of the tumour eQTLs are exclusive to CRC and show that this specificity is partially driven by increased expression of specific transcription factors and changes in methylation patterns. We show that tumour-specific eQTLs are more enriched for low CRC genome-wide association study (GWAS) *P* values than shared eQTLs, which suggests that some of the GWAS variants are tumour specific regulatory variants. Importantly, tumour-specific eQTL genes also accumulate more somatic mutations when compared to the shared eQTL genes, raising the possibility that they constitute germline-derived cancer regulatory drivers. Collectively the integration of genome and the transcriptome reveals a substantial number of putative somatic and germline *cis*-regulatory cancer changes that may have a role in tumorigenesis.

The non-coding genome has so far been overlooked in the search for drivers in cancer, except for some isolated examples^{1,2}. The genome contains regulatory non-coding germline variation affecting gene expression, namely eQTLs, which are major components in complex disease predisposition³, and these have not been examined within the context of tumorigenesis. Epistasis between eQTLs and coding mutations in genes has a role in disease⁴; therefore it is likely that interactions between somatic regulatory variants or eQTLs and coding variation are important in tumorigenesis. To examine this we performed an RNA sequencing analysis of 103 matched tumour and normal colon mucosa CRC samples (Supplementary Fig. 2) of the SYSCOL consortium (Supplementary Table 1). Ninety samples were genotyped for their germline genome and imputed to 1000 Genomes phase 1 release⁵ (Supplementary Methods). We also RNA-sequenced 20 reference tissues to create a more comprehensive control transcriptome for CRC. A general overview and main findings are summarized in Supplementary Fig. 3.

We examined transcriptome perturbations during tumorigenesis by identifying differentially expressed genes (DEGs) and find 1,676 DEGs (false discovery rate (FDR) = 5%, fold change ≥ 2) (Supplementary Table 2 and Supplementary Fig. 4). The functional terms enriched in DEGs are given in Supplementary Table 3, and pathways impacted, which include known CRC pathways, are given in Supplementary Table 4. There

is no significant difference of differential expression between known CRC driver genes^{6,7} or The Cancer Genome Atlas (TCGA) pan-cancer genes^{7,8}, and expression matched random genes (Supplementary Fig. 5). In addition, we find 213 differentially spliced genes (FDR = 5%) (Supplementary Table 5, Supplementary Figs 6 and 7). Among these are *TGFBI* (ref. 9), *RASA1* (ref. 10), *SLC26A3* (ref. 11) and *SLC39A14* (ref. 12), which have been previously implicated in CRC. The enriched functional terms are given in Supplementary Table 6.

We tested the relationship of samples' gene expression and searched for regions of the genome in which CRC modifies the correlation of proximal gene expression. We observe that the tumour samples, normal samples, and the 20 reference tissue samples form three distinct clusters (Supplementary Fig. 8), and the variance of the CRC transcriptome is increased when compared to the normal samples (Supplementary Fig. 9). In all chromosomes there are blocks of proximal genes whose correlations are altered in CRC (Supplementary Figs 10 and 11, and Supplementary Table 7) pointing to coordinated regulation likely due to consistent epigenetic effects and locus control regions. The results above suggest a rewiring of the regulatory networks in CRC.

To further dissect *cis*-regulatory effects in CRC we examined variation in regulation of gene expression. One powerful method for discovering *cis*-regulatory variability is ASE analysis, which was conducted as described previously¹³. We find that approximately 10% of coding heterozygous sites exhibit significant ASE (FDR = 1%, $P < 0.001$). The proportion of sites with ASE is significantly higher in tumours compared to normal samples (Fig. 1a) and approximately 34% of ASE is CRC-specific. This excess is likely to be an indication of copy number alterations (CNAs) that have been described previously^{6,14,15}. Furthermore, in approximately 10% of ASE sites significant in both matched samples, the direction of the ASE effect is reversed (Fig. 1b), also indicating genes found in potential CNA regions or genes where the *cis*-regulatory landscape has been altered.

We propose that the somatic events in the regulatory regions of genes are reflected in the difference of allelic ratios between matched pairs at expressed heterozygous sites. We define ASE somatic events (Supplementary Methods) and observe a significant correlation (Spearman's $\rho = 0.05$, $P = 8.8 \times 10^{-7}$) between ASE somatic event rates and coding somatic mutation rates¹⁶ of genes, indicating that somatic ASE events may be more likely to be selected in genes implicated in CRC (Fig. 2a and Supplementary Fig. 12). To determine a score of somatic dysregulation we compared the ASE somatic event rate of each gene to all other genes (Supplementary Methods). The score is defined as the enrichment of low *P* values (π_1)¹⁷. We observe a bimodal distribution of π_1 values where most genes are similar to other genes but a small fraction of genes demonstrate high π_1 , representing genes with dysregulation (Supplementary Fig. 13 and Supplementary Table 8). Known CRC driver genes^{6,7} and TCGA pan-cancer genes^{7,8} have significantly higher π_1 compared to other genes (Supplementary Fig. 14). Therefore, the π_1

¹Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland. ²Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, 1211 Geneva, Switzerland. ³Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland. ⁴Department of Molecular Medicine, Aarhus University Hospital, 8000 Aarhus, Denmark. ⁵Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 Barcelona, Catalonia, Spain. ⁶Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. ⁷Nuffield Department of Clinical Medicine and Oxford NIHR Comprehensive Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ⁸Department of Physiological Sciences II, School of Medicine, University of Barcelona, 08007 Barcelona, Barcelona, Spain. ⁹Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

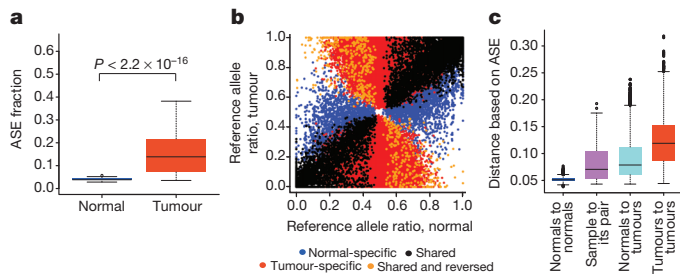


Figure 1 | Allele-specific expression. **a**, Boxplots showing the distribution of the proportion of heterozygous sites with significant ($P < 0.001$) ASE in normal and tumour samples. This is significantly higher in tumours, which is suggestive of CNA. **b**, Allele ratios of significant ($P < 0.001$) ASE sites in tumours versus matched normals. Approximately 34% of the ASE sites are tumour-specific (red points) and approximately 10% of the shared ASE sites (black and orange points) show reversal of the effect (orange points). This is indicative of CNA or *cis*-regulatory changes in these samples. **c**, Boxplots showing the distributions of pairwise distances between pairs of samples calculated from ASE ratios. Allelic expression is more similar between tumours and their matched controls (magenta) than other comparisons indicating that germline genotypes remain important determinants of gene expression even after tumorigenesis. All distributions are significantly different from each other (Mann–Whitney U -test, $P < 2.2 \times 10^{-16}$). In the boxplots the black horizontal line represents the median, the boxes are delimited by versions of the first and the third quartile, and whiskers extend to 1.5 times the box length with points outside of these represented as circles.

score contains information about the driving capability of *cis*-regulatory somatic mutations in highly scoring genes.

To define genes with an excess of somatic *cis*-regulatory events we used principles similar to those used for identifying excess of protein-coding somatic events (Supplementary Methods). We detected 71 genes (FDR = 5%) with significantly higher ASE somatic event rates, which we define as genes with allelic dysregulation (GADs) (Supplementary Methods, Fig. 2c and Supplementary Table 8), which also have high π_1 (Fig. 2b). These GADs are significantly enriched for TCGA pan-cancer drivers^{7,8} and known CRC driver genes^{6,7} (seven pan-cancer genes $P = 0.035$, two CRC genes $P = 0.039$; Supplementary Methods), indicating that we are capturing known cancer genes. This suggests two scenarios—somatic coding mutations and regulatory mutations undergo epistatic

selection, or genes become involved in tumorigenesis in the presence of either—and we believe that both are likely explanations. We tested whether GADs can be driven by overexpression of genes in tumours, and observe that GADs are not significantly more overexpressed than other genes (Mann–Whitney U -test, $P = 0.1$; Supplementary Fig. 15). We assessed whether systematic CNAs were responsible for GADs (Supplementary Methods). There is no significant clustering of GADs (Supplementary Fig. 16) and although 25% of the ASE somatic events in GADs overlap with CNAs, this is lower than non-GADs (30%) and is not significantly different between GADs and non-GADs (Supplementary Fig. 17). These two results together suggest that the contribution of CNAs to the identification of GADs is not significant. Our ASE methodology, while taking into account many of the known biases, remains imperfect. However, this does not have significant influence in our analysis, as the majority of biases are expected to be shared between normal samples and tumours. Therefore by using recurrent ASE differences in matched tumour and normal samples at germline sites as a proxy to the changes in the somatic regulation of genes, we have determined a set of genes with putative *cis*-regulatory driving mutations.

To assess the maintenance of germline *cis*-regulatory effects between normal and cancer tissue we looked at patterns of ASE. We calculated ASE distance of tumour and normal samples to normal and tumour samples in the same individual and other individuals (Supplementary Fig. 18). This shows that a tumour sample is most similar to its matched normal colon mucosa sample (Fig. 1c), meaning that allelic expression of most genes is conserved after tumorigenesis. Thus the germline genotypes remain key determinants of allelic gene expression in tumours.

One of the open questions in tumorigenesis is whether non-coding germline variants contribute as driving factors¹⁸. We propose that if such variants existed they would affect gene expression specifically in tumours. To address this question we conducted a *cis*-eQTL analysis (Supplementary Methods). We find 1,693 and 948 eQTLs (permutation $P < 0.01$, FDR 9.4 and 16.1) in normal samples and tumours, respectively (Supplementary Tables 9 and 10). At this threshold, 61% of the tumour eQTLs and 78% of normal eQTLs appear to be tissue specific, with 368 shared eQTLs (Fig. 3c). Using a more sensitive approach¹⁷ we estimate that 64% of tumour eQTLs and 62% of the normal eQTLs are shared (Fig. 3d). We find stronger eQTL effects and more sharing closer to the transcription start site (TSS) (Fig. 3a). The effect sizes of eQTLs

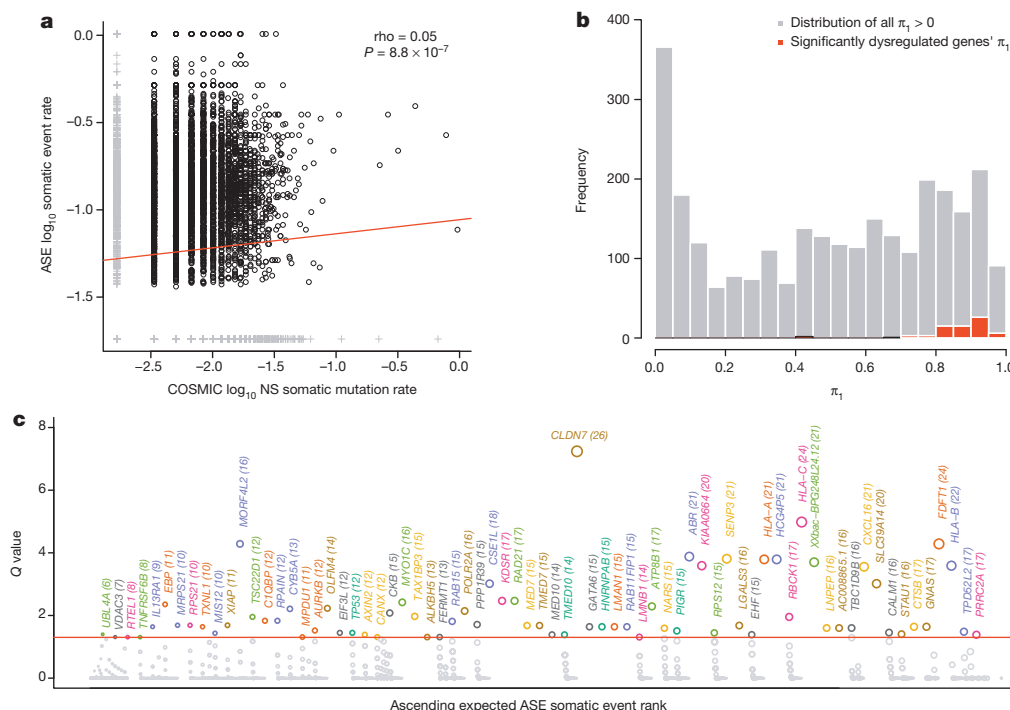


Figure 2 | Genes that are significantly dysregulated in CRC. **a**, The correlation of ASE somatic event rate and non-synonymous (NS) coding somatic mutation rates of genes (the grey crosses are zeros that were transformed to the minimum non-zero value). **b**, Distribution of the π_1 score for dysregulation. High π_1 scores indicate genes that are dysregulated in CRC. Grey, distribution for all genes with a $\pi_1 > 0$; red, distribution for the GADs (see c). **c**, 71 GADs that accumulate significantly more ASE somatic events than expected by chance. The genes are ordered by their expected ASE somatic event count on the x axis (GADs are equidistant from each other). The y axis is the Benjamini–Hochberg corrected $-\log_{10} P$ value. The size of the circle is proportional to the ASE somatic event count, which is given in parenthesis after each gene's name. The red horizontal line corresponds to the 5% FDR cut-off.

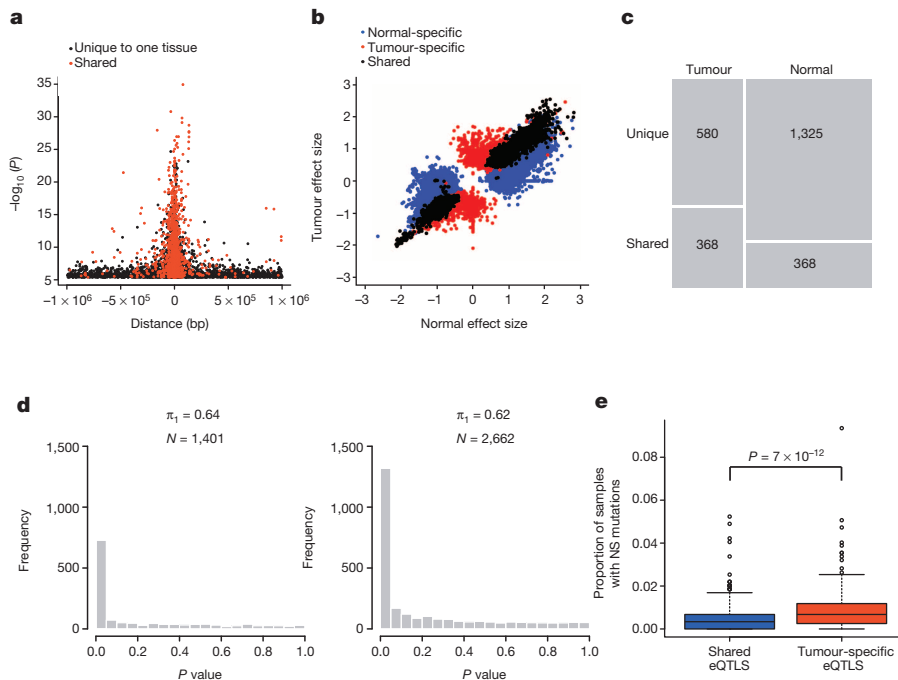


Figure 3 | *cis*-eQTLs. **a**, Tissue specificity and distance of eQTL to transcript start site (TSS). The shared eQTLs (red) are closer to the TSS than are the tissue specific eQTLs (black) (Spearman's $\rho = -0.28$, $P < 2.2 \times 10^{-16}$) and distance to TSS and significance are correlated (Spearman's $\rho = -0.44$, $P < 2. \times 10^{-16}$). **b**, Effect sizes of eQTLs. The direction of the effect is conserved in the shared eQTLs meaning that the germline genotypes are the main drivers in tumour eQTLs. **c**, Mosaic plot of tissue specificity of eQTLs. Of eQTLs, 61% and 78% are specific to the tumours and to the normal samples, respectively. **d**, P-value distributions of significant SNP-exon pairs tested in the other tissue. The π_1 statistic estimates the tissue sharing of eQTLs. **e**, Distributions of the proportion of large intestine samples in the COSMIC database with a confirmed protein altering somatic mutation for the shared and tumour specific eQTL genes. Tumour specific eQTL genes accumulate significantly more somatic mutations (Mann-Whitney *U*-test $P = 7 \times 10^{-12}$) making some of these likely germline regulatory drivers. In the boxplots the black horizontal line represents the median, the boxes are delimited by versions of the first and the third quartile, and whiskers extend to 1.5 times the box length with points outside of these represented as circles.

show that the direction of shared eQTL effects is conserved after tumorigenesis (Fig. 3b), thus we expect the tumour-specific eQTLs to be similarly driven by germline genotypes. A strict set of tumour-specific eQTLs was defined (Supplementary Methods) and we assessed the differential expression between shared eQTLs and tumour-specific eQTLs. Although there

is a very small (median 1.05-fold) but significant (Mann-Whitney *U*-test, $P = 0.02$; Supplementary Fig. 19) increase in expression of tumour-specific eQTL genes, it is not sufficient to explain the tumour-specific eQTLs. Using an interaction model with tumour stage we assessed whether behaviour of the tumour-specific eQTLs change as cancer progresses.

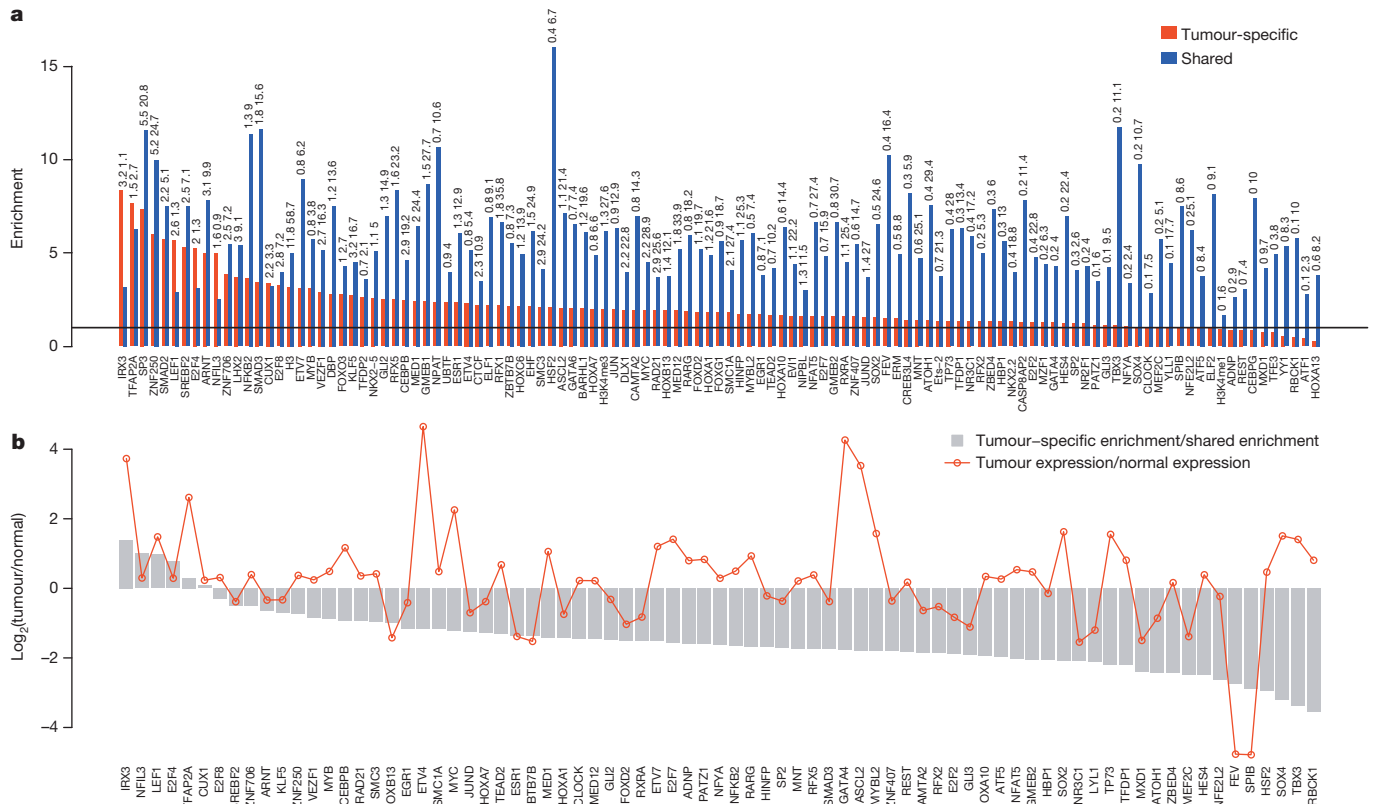


Figure 4 | Functional enrichments of eQTLs. **a**, The bar plot is ordered by the tumour specific enrichment. The null (frequency and distance matched) is represented as the black horizontal line. The numbers above each bar are the $-\log_{10} P$ values of the enrichment, tumour-specific first followed by the shared. **b**, The ratio of the tumour specific enrichment to the shared enrichment (in log

scale) for significantly differentially expressed transcription factors are shown as grey bars. The red line depicts the differential expression fold change of the transcription factors. The first six transcription factors (IRX3, E2F4, NFIL3, TFAP2A, CUX1 and LEF1), where we have higher enrichments in the tumour specific eQTLs, are also upregulated in CRC.

Although no results survive multiple test correction, the top hit is *THBS3* (nominal $P = 0.0007$), which is a stimulator of tumour progression in osteosarcoma¹⁹, making it a plausible candidate for a stage-regulated eQTL.

To assess whether the eQTL variants were biologically functional, we used the Ensembl Regulatory Build to look for enrichment of eQTL variants in non-coding functional regions (Supplementary Methods). We find significant enrichments for many marks (Supplementary Fig. 20), highlighting the functional relevance of the variants discovered. To understand whether different types of functional effects were driving tumour-specific eQTLs and shared eQTLs we used the ChIP-seq (chromatin immunoprecipitation followed by sequencing) peaks from the colon cancer LoVo cell line²⁰ (Fig. 4a). We observe binding sites of six transcription factors that have stronger enrichment in the tumour-specific eQTLs. All of these transcription factors are also significantly upregulated in CRC (Fig. 4b and Supplementary Table 11) and there is a significant positive correlation between the tumour-specific eQTL enrichment to shared eQTL enrichment ratio and fold change in the expression of the corresponding transcription factors ($r = 0.24$, $P = 0.01$; Supplementary Fig. 21), indicating that differential expression of these factors are likely to be driving a proportion of the tumour-specific eQTLs. We also assessed methylation patterns and find that there is a significant increase in differential methylation in tumour-specific eQTL variant regions compared to shared eQTLs. (Mann–Whitney U -test, $P = 2.5 \times 10^{-7}$, Supplementary Methods and Supplementary Fig. 22). The changes in methylation indicate regulatory switches responsible for some of the tumour-specific eQTL effects. We also estimate¹⁷ that up to 38% tumour-specific eQTLs are active in at least another healthy tissue (Supplementary Methods and Supplementary Table 12). Finally, we tested for enrichment of low P values in CRC GWAS amongst the eQTLs (Supplementary Methods). The highest levels of enrichment for low GWAS P values is seen in tumour-specific eQTLs (π_1 of 11% versus 7% in shared eQTLs; Supplementary Fig. 23), signifying that a proportion of the CRC GWAS signals are *cis*-regulatory variants active only in the tumours.

To test whether the tumour-specific eQTLs are likely to be drivers in CRC, we proposed that the *cis*-regulatory changes may have similar impact to somatic mutations in genes with the same or similar function; therefore we would expect to see an increased somatic mutation rate in the tumour-specific eQTL genes. We compared the proportion of CRC samples in the COSMIC¹⁶ database that had a protein altering somatic mutation between the 376 tumour-specific eQTL genes and the 368 shared eQTL genes, and find that tumour-specific eQTL genes accumulate significantly more somatic mutations (Mann–Whitney U -test, $P = 7 \times 10^{-12}$, Fig. 3e and Supplementary Table 13). Moreover, there is a significant 2.5-fold enrichment (Fisher's exact test, $P = 0.02$) for TCGA pan-cancer driver genes^{7,8} in tumour-specific eQTL genes (20 versus 8 in shared eQTL genes), making these likely non-coding *cis*-regulatory drivers in CRC. To avoid any potential detection bias we tested whether the tumour-specific eQTL genes also accumulate more mutational events by comparing inferred CNA in shared eQTL genes versus tumour-specific eQTLs using ASE estimates. There is no significant increase in inferred CNA of tumour-specific eQTLs (Fisher's exact test, $P = 0.5$; Supplementary Fig. 24). Collectively these results reveal germline variants, whose functions are activated under the tumour state, some of which are likely to be selected in tumour progression.

Here we present the allelic transcriptome changes that occur in CRC tumorigenesis. We discover 71 GADs and 376 genes with tumour specific germline *cis*-regulatory variants. Both categories demonstrate characteristics that support their role as putative cancer drivers. This gives us access to putative non-coding somatic and germline CRC drivers on an unprecedented scale. In addition, tumour specific *cis*-eQTLs reveal a new category of variants that are likely to contribute to cancer besides predisposing alleles and somatic mutations. It is likely that some of the predisposing variants discovered via GWAS are in fact germline drivers. We demonstrate that integration of genome and transcriptome followed by ASE and eQTL analysis in normal–tumour matched samples can be used to identify important non-coding regulatory effects in cancer.

Received 19 December 2013; accepted 19 June 2014.

Published online 23 July 2014.

- Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
- Dvinge, H. *et al.* The shaping and functional consequences of the microRNA landscape in breast cancer. *Nature* **497**, 378–382 (2013).
- Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nature Genet.* **44**, 1084–1089 (2012).
- Lappalainen, T., Montgomery, S. B., Nica, A. C. & Dermitzakis, E. T. Epistatic selection between coding and regulatory variation in human evolution and disease. *Am. J. Hum. Genet.* **89**, 459–463 (2011).
- The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- The Cancer Genome Atlas Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
- Perez-Llamas, C. & Lopez-Bigas, N. Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE* **6**, e19541 (2011).
- Tamborero, D. *et al.* Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Scientific Rep.* **2**, 2650 (2013).
- Ma, C. *et al.* Extracellular matrix protein β ig-h3/TGFB1 promotes metastasis of colon cancer by enhancing cell extravasation. *Genes Dev.* **22**, 308–321 (2008).
- Sun, D. *et al.* MicroRNA-31 activates the RAS pathway and functions as an oncogenic MicroRNA in human colorectal cancer by repressing RAS p21 GTPase activating protein 1 (RASAP1). *J. Biol. Chem.* **288**, 9508–9518 (2013).
- Schweinest, C. W., Henderson, K. W., Suster, S., Kondoh, N. & Papas, T. S. Identification of a colon mucosa gene that is down-regulated in colon adenomas and adenocarcinomas. *Proc. Natl Acad. Sci. USA* **90**, 4166–4170 (1993).
- Thorsen, K. *et al.* Alternative splicing of SLC39A14 in colorectal cancer is regulated by the Wnt pathway. *Mol. Cell. Proteom.* **10**, M110.002998 (2011).
- Lappalainen, T. *et al.* Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
- Emi, M. *et al.* Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. *Cancer Res.* **52**, 5368–5372 (1992).
- Matsuzaki, K. *et al.* The relationship between global methylation level, loss of heterozygosity, and microsatellite instability in sporadic colorectal cancer. *Clin. Cancer Res.* **11**, 8564–8569 (2005).
- Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
- Nica, A. C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
- Aerts, S. & Cools, J. Cancer: Mutations close in on gene regulation. *Nature* **499**, 35–36 (2013).
- Dalla-Torre, C. A. *et al.* Effects of *THBS3*, *SPARC* and *SPPI* expression on biological behavior and survival in patients with osteosarcoma. *BMC Cancer* **6**, 237 (2006).
- Yan, J. *et al.* Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell* **154**, 801–813 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was designed under the SYSCOL project and samples were collected with informed consent in accordance with local law. This research is supported by grants from European Commission SYSCOL FP7 (UE7-SYSCOL-258236), European Research Council (ERC 260927), Louis Jeantet Foundation, Swiss National Science Foundation (130326, 130342), the NIH-NIMH (MH090941), the Danish National Advanced Technology Foundation, the John and Birthe Meyer Foundation, the Danish Council for Independent Research (Medical Sciences), the Danish Council for Strategic Research, the Danish Cancer Society, The Cellex Foundation, the Botin Foundation, the Fundacion Olga Torres, and the Health and Science Departments of the Catalan Government (Generalitat de Catalunya). The Danish Cancer Biobank is acknowledged for biological material. We thank S. Moran, D. Garcia and C. Arribas for their technical support. This study was also funded by Cancer Research UK and The Oxford Comprehensive Biomedical Research Centre (I.P.T.). Core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford, was provided by grant (090532/Z/09/Z). Cancer Research UK provided funding individually to R.S.H. (C1298/A8362–Bobby Moore Fund for Cancer Research UK). This study made use of genotyping data from the 1958 Birth Cohort and NBS samples, kindly made available by the Wellcome Trust Case Control Consortium 2. The computations were performed at the Vital-IT (<http://www.vital-it.ch>) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

Author Contributions H.O., C.L.A., J.B.B., T.F.O. and E.T.D. designed the study. H.O. and E.T.D. coordinated the project. H.O., J.B.B., A.P., I.P., D.B., L.R. and M.H.R. participated in RNA-sequencing data production. J.S., E.V. and M.E. designed and conducted the methylation experiment. N.W., I.T. and R.S.H. conducted the CRC GWAS. H.O. and P.G.F. analysed the data. H.O., C.L.A. and E.T.D. drafted the paper.

Author Information The RNA-sequencing and genotype data are deposited in the European Genome-phenome Archive (EGA, <https://www.ebi.ac.uk/ega/>) for controlled access; the study accession number is EGAS00001000854. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.O. (halit.ongen@unige.ch) or E.T.D. (emmanouil.dermizakis@unige.ch).

Genome-scale functional characterization of *Drosophila* developmental enhancers *in vivo*

Evgeny Z. Kvon^{1†}, Tomas Kazmar¹, Gerald Stampfel^{1*}, J. Omar Yáñez-Cuna^{1*}, Michaela Pagani¹, Katharina Schernhuber¹, Barry J. Dickson^{1†} & Alexander Stark¹

Transcriptional enhancers are crucial regulators of gene expression and animal development¹ and the characterization of their genomic organization, spatiotemporal activities and sequence properties is a key goal in modern biology^{2–8}. Here we characterize the *in vivo* activity of 7,705 *Drosophila melanogaster* enhancer candidates covering 13.5% of the non-coding non-repetitive genome throughout embryogenesis. 3,557 (46%) candidates are active, suggesting a high density with 50,000 to 100,000 developmental enhancers genome-wide. The vast majority of enhancers display specific spatial patterns that are highly dynamic during development. Most appear to regulate their neighbouring genes, suggesting that the *cis*-regulatory genome is organized locally into domains, which are supported by chromosomal domains, insulator binding and genome evolution. However, 12 to 21 per cent of enhancers appear to skip non-expressed neighbours and regulate a more distal gene. Finally, we computationally identify *cis*-regulatory motifs that are predictive and required for enhancer activity, as we validate experimentally. This work provides global insights into the organization of an animal regulatory genome and the make-up of enhancer sequences and confirms and generalizes principles from previous studies^{1,9}. All enhancer patterns are annotated manually with a controlled vocabulary and all results are available through a web interface (<http://enhancers.starklab.org>), including the raw images of all microscopy slides for manual inspection at arbitrary zoom levels.

Animal development depends on differential gene expression governed by genomic regulatory elements called enhancers^{1,9}, which are being studied extensively^{2,3,8,10,11}. Many of the basic principles of developmental gene regulation have been elucidated in the fruitfly *Drosophila melanogaster*^{1,12,13}, and work over the past decades has characterized gene expression, transcription factor binding, chromatin features and enhancer activity in *Drosophila* at unprecedented levels^{2,5–8,14–16}. This and the ability to obtain many embryos from all developmental stages¹⁷ make *Drosophila* an ideal model in which to characterize spatiotemporal enhancer activities at a genomic scale and throughout embryogenesis.

To systematically characterize developmental enhancers in the *D. melanogaster* genome, we made use of transgenic fly lines (Vienna Tiles (VT) library), publicly available from the Vienna *Drosophila* RNAi Center (VDRC). Each line contains a transcriptional reporter construct with a ~2 kilobase (kb) genomic DNA fragment (enhancer candidate), minimal promoter and *GAL4* reporter gene integrated into an identical position in the fly genome¹⁶, thus allowing the direct comparison of the candidates' activities (Fig. 1a, Extended Data Fig. 1a, b and Supplementary Table 1). Together, these fragments cover about 14 million base pairs or 13.5% of the non-coding, non-repetitive genome, with little or no bias regarding the distance to transcription start sites (TSSs; Extended Data Fig. 1c) or the embryonic expression of neighbouring genes (Extended Data Fig. 1d).

We developed a high-throughput pipeline to assess transcriptional enhancer activities in fly embryos by *in situ* hybridization against the *GAL4* reporter transcript. For each transgenic line, we acquired whole-slide

images, each with about 400 embryos covering all stages of embryogenesis, and manually annotated the enhancer activity patterns using a controlled vocabulary¹⁴ at six time intervals of embryogenesis (Extended Data Fig. 1e). The pipeline reported activities independent of fragment delineation and orientation and recovered 27 out of 28 known enhancers, whereas 13 out of 13 non-*Drosophila* controls were inactive (Extended Data Fig. 2a–c and Supplementary Information section 1). Results from re-testing 34 negative and 78 positive fragments using a different genomic site (on chromosome 2L instead of 3L) and reporter gene (*lexA*) were highly similar and the majority (82%) of enhancer activity patterns matched to the expression patterns of neighbouring genes, suggesting that we predominantly measured endogenous enhancer activities (Extended Data Fig. 2d–f and Supplementary Information section 1).

3,557 of all 7,705 tested candidate fragments (46%) were active in the embryo with diverse patterns that included gap and pair-rule patterns, all primary germ layers (Extended Data Figs 3a and 4a), and all major cell types and tissues (Fig. 1b and Extended Data Figs 3b and 4b). The fraction of active fragments increased about fivefold from ~7% in early embryos (stages 4–6) to ~35% for stages 15–16, consistent with the increase in organism complexity and the number of cell types (Fig. 1c). By contrast, the number of expressed genes remains roughly constant during embryogenesis (~1.3-fold increase¹⁸). Enhancer activities were much sparser than gene expression patterns both temporally and spatially: while 94% of all enhancers were only transiently active and only 0.8% were ubiquitous during the entire embryogenesis, this was true for 56.7% and 20.5% of the genes, respectively (Fig. 1d, e and Extended Data Fig. 5a–c). The temporal dynamics of enhancer activity was also apparent from changes of enhancer-associated chromatin features such as DNase I hypersensitivity (DHS), binding of co-activator CBP/p300, and presence of histone H3K27 acetylation mark assessed in entire embryos or in a tissue-specific manner^{2,19,20} (Fig. 1f–h, Extended Data Fig. 6 and Supplementary Information section 2). Together, this confirms and quantifies the transient and dynamic nature of enhancer function and suggests that development progresses through increasingly complex gene regulation by enhancers with temporally and spatially restricted activities.

We next identified domains in the blastoderm embryo in which enhancers appeared co-regulated (that is, were coordinately active or inactive). Automated image segmentation and reverse clustering revealed distinct regions corresponding to the presumptive anterior and posterior endoderm, head and trunk mesoderm, procephalic neuroectoderm, and others, overall strongly resembling the established fate map of the blastoderm embryo¹⁷ (Fig. 1i and Extended Data Fig. 4c). This suggests that cells within these domains have a common developmental fate, presumably due to shared *trans*-regulatory environments. Indeed, during late stages, early mesodermal enhancers were preferentially active in mesoderm derivatives (somatic, visceral and cardiac muscles), whereas early endodermal enhancers were active in endoderm derivatives (midgut and Malpighian tubules) (Extended Data Fig. 4d). These and equivalent trends for other presumptive tissues of the early embryo (Extended Data Fig. 4e–g)

¹Research Institute of Molecular Pathology (IMP), Vienna Biocenter VBC, Dr Bohr-Gasse 7, 1030 Vienna, Austria. [†]Present addresses: Howard Hughes Medical Institute, Janelia Farm Research Campus, Ashburn, Virginia 20147, USA (B.J.D.); Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA (E.Z.K.).

*These authors contributed equally to this work.

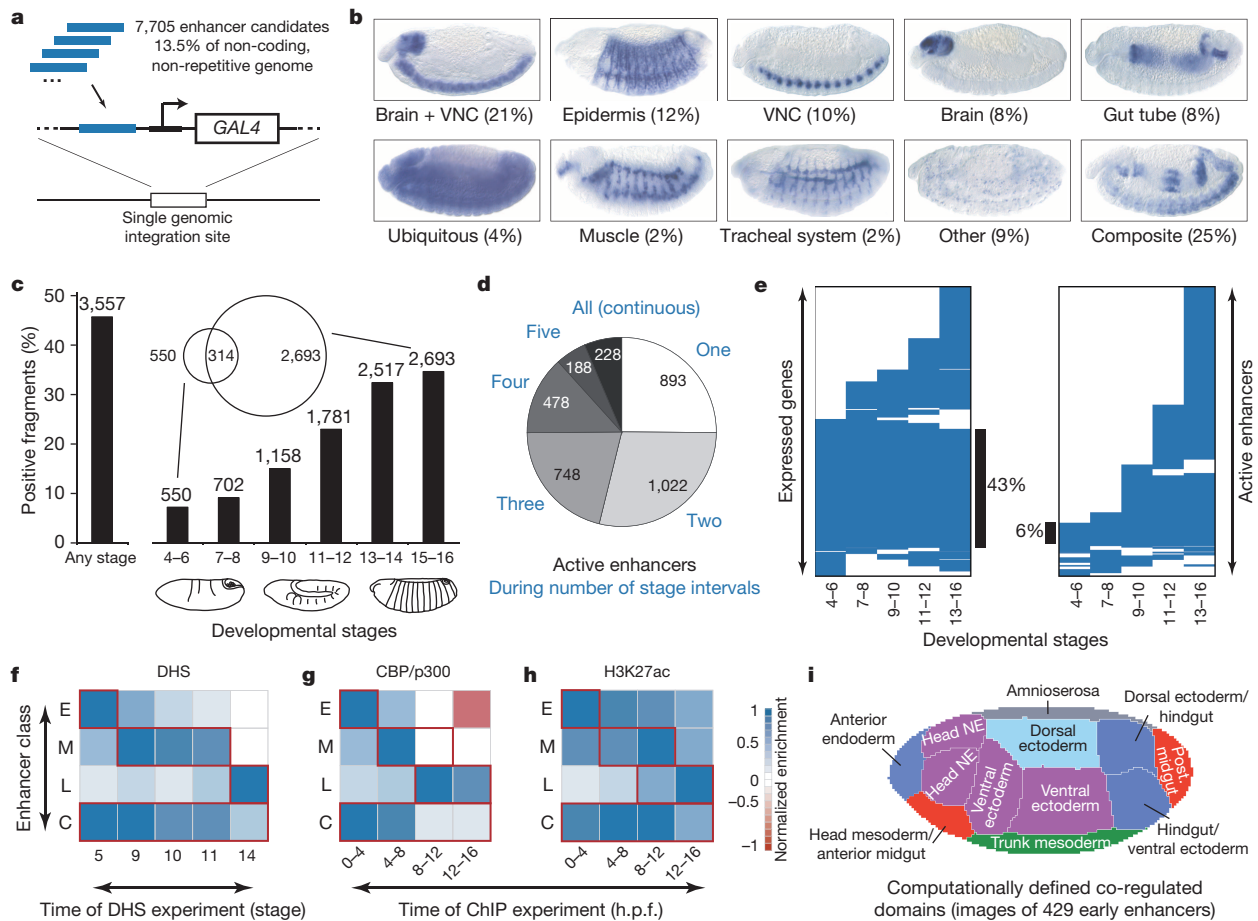


Figure 1 | Enhancers display highly diverse and dynamic activity patterns across *Drosophila* development. **a**, The VT library comprises transgenic flies with candidate fragments (blue) upstream of a transcriptional reporter (middle) in a constant genomic landing site (Extended Data Fig. 1). **b**, Proportion of enhancer activities in prominent tissues at stages 13–14 (representative embryos; Extended Data Figs 3 and 4). VNC, ventral nerve cord. **c**, The number of active enhancers increases during embryogenesis, with some overlap between early and late enhancers (Venn diagram in **c**). **d**, 3,329 (94%) embryonic enhancers are only transiently active. **e**, Temporal dynamics of gene expression (left, 5,134 genes¹⁴) and enhancer activity (right, 3,557

demonstrate that enhancer activities are consistent with the progression of development along defined cell lineages, highlighting the gene regulatory basis of development.

To analyse the locations of enhancers with respect to their putative target genes, we assigned enhancers to genes by manually matching enhancer activity and gene expression patterns (Figs 2a and 3a). For the 874 enhancers with the strongest activity patterns, we considered 3,681 genes within five genes up- and downstream of each enhancer (including host genes for intronic enhancers), that is, 9,293 enhancer–gene pairs. For 4,224 of these pairs (45%; 1,690 genes), expression patterns were available, resulting in 482 enhancer-to-gene assignments (of the enhancers for which all neighbouring genes were characterized, 82% could be assigned; Extended Data Fig. 2f and Supplementary Table 4). The assignments were supported by the location of chromosomal domain boundaries²¹, binding sites of insulator proteins²² and evolutionary chromosome breakpoints²³, all of which were depleted between enhancers and their assigned targets (Fig. 2b–d, Extended Data Fig. 7a–c and Supplementary Information section 3). Twenty-eight enhancers were assigned to and potentially regulate two genes, 23 of which were paralogues with very similar expression patterns (Supplementary Information section 4). During stages 4–6, 16 genes were assigned to enhancers with overlapping or identical activities reminiscent of shadow enhancers²⁴. This is a considerable fraction

enhancers; black vertical lines indicate continuous expression or activity; Extended Data Fig. 5a). **f–h**, Heatmaps show the median enrichment of DNA accessibility²⁰ (**f**), CBP/p300 binding (**g**) and H3K27 acetylation (ac) marks² (**h**) on early (E), middle (M), late (L) and continuous (C) enhancers (rows) for experiments performed at different time points during *Drosophila* development (columns; red highlights coinciding time points; Extended Data Fig. 6). **i**, Co-regulated domains defined by reverse clustering of raw image data for 429 early enhancers resemble the embryo fate map¹⁷ (Extended Data Fig. 4c). ChIP, chromatin immunoprecipitation; NE, neuroectoderm.

(14%) among all 116 genes with multiple enhancers, in particular for developmental regulators (14 out of 16 genes are transcription factors; Supplementary Information section 5).

Along the linear genomic DNA sequence, the distances between the enhancers and the TSSs of their assigned target genes varied greatly: although many such pairs were close (21% were <4 kb), the median distance was 10 kb, and 28% of all inferred regulatory interactions were distal (>20 kb), up to more than 100 kb (Fig. 2e). However, considering the location of genes, the vast majority (88%) of all enhancers were located in the vicinity of their targets (Fig. 2f). Nevertheless, 12% of all enhancers were assigned across intervening genes and appeared to skip one (8%) or more (4%) genes to regulate a distal gene (Fig. 2f), as found for a *Sex combs reduced* (*Scr*) enhancer that lies beyond the *fushi tarazu* (*ftz*) gene²⁵. Interestingly, enhancers were located almost as frequently upstream (30%) as downstream (22%) of their target genes (for example, the *SoxNeuro* (*SoxN*) locus; Extended Data Fig. 8), suggesting that no particularly preferred relative enhancer location might exist.

Thirty-six per cent of the enhancers were intragenic and appeared to predominantly (79%) regulate their host genes, as exemplified by *Thrombospondin* (*Tsp*; Fig. 3a). However, 21% were assigned to a neighbouring gene instead (Fig. 3b), including an enhancer located inside the intron of *bric a brac 1* (*bab1*) that appears to activate *bab2* over a distance of

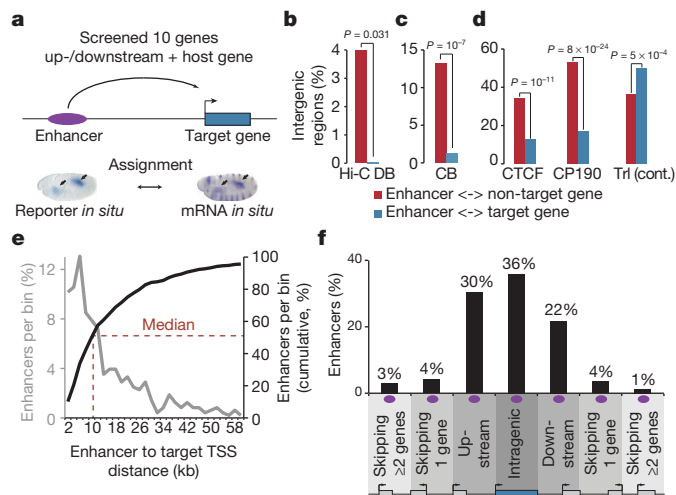


Figure 2 | The organization of the *Drosophila* cis-regulatory genome.

a, Enhancer to target gene assignment based on enhancer activity and gene expression patterns. **b–d**, Chromosomal domain boundaries determined by Hi-C²¹ (**b**), breakpoints during genome evolution²³ (**c**) and insulator binding sites²² (**d**) show relative depletions between enhancers and their assigned target genes (blue) and enrichments between enhancers and non-targets (red), whereas the opposite is true for the activator Trl²² (binomial P values are shown; see Extended Data Fig. 7a–c for additional insulator proteins and details). **e**, Genomic distances between enhancers and their assigned target gene TSSs in kb (grey, frequencies; black, cumulative). **f**, Frequency of enhancers (purple) at different genomic positions relative to their target genes (blue; schematic locus). Eighty-eight per cent of all enhancers are in the genes' genomic neighbourhoods within regulatory domains. CB, chromosomal breakpoints; DB, domain boundaries.

93 kb (Fig. 3c). *bab1* is not detectably expressed in the embryo during the corresponding developmental stage, which we found to be true more generally when intragenic enhancers regulated flanking genes rather than their host genes (Fig. 3d). Similarly, when intergenic enhancers were assigned to distal genes, the skipped genes were significantly less highly expressed than the target genes (Fig. 3d). Together these results support a predominantly local organization of the *Drosophila* genome into regulatory domains reminiscent of the chromosomal domains inferred from chromatin interactions²¹.

The agreement of most enhancers' activities with the expression patterns of neighbouring genes (Figs 2f, 3a and Extended Data Figs 2f, 8) confirms that enhancer activity is predominantly context independent⁹. However, 18% of the enhancers could not be assigned to neighbouring genes (Extended Data Fig. 2f) and might be involved in more distal regulation (for example, ref. 26). For 19%, the activities were similar but broader and might thus be modulated in the endogenous sequence contexts in a more complex fashion (Supplementary Information section 1). Such context dependence is known for several loci in *Drosophila* (for example, the Hox locus²⁷) and mouse (for example, *Fgf8* (ref. 28)), and enhancers in the bithorax complex indeed matched to gene expression patterns during early stages but appeared broader later (Extended Data Fig. 7d).

Many different enhancers showed similar or identical activity patterns in various embryonic tissues. For example, 263 were active throughout the central nervous system (CNS), 59 in midgut and 32 in macrophages (Extended Data Fig. 3), thus probably providing sufficient statistical power to discern predictive sequence signatures. Indeed, the motif content alone allowed the discrimination of enhancers from different functional classes using supervised machine learning in a cross-validated setting²⁹ (Extended Data Fig. 9a, b and Supplementary Table 5). The

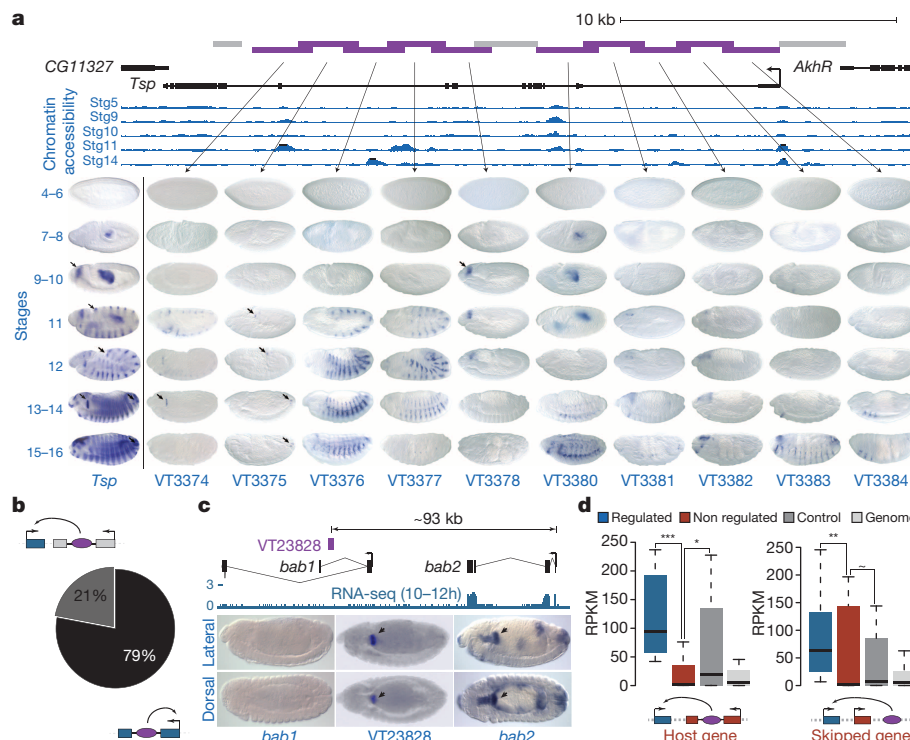


Figure 3 | Intragenic enhancers in the *Drosophila* genome. **a**, Enhancers in the *Tsp* locus. Top, UCSC Genome Browser screenshot including tested fragments (purple, positive; grey, negative) and DNA accessibility²⁰. Bottom, embryos for all six time points of embryogenesis (left, *in situ* visualizing *Tsp* mRNA¹⁴; arrows highlight small expression/activity domains). **b**, Twenty-one per cent of intragenic enhancers are assigned to a neighbouring gene. **c**, A distal *bab2* enhancer (VT23828) in the intron of a neighbouring gene *bab1*. Top, UCSC Genome Browser screenshot including RNA-seq data for the

corresponding stages¹⁸. Bottom, embryo images depicting the *bab1* and *bab2* expression during stages 13–14 (ref. 14) and VT23828's activity in the proventriculus (middle). **d**, Non-regulated host and skipped genes are often not expressed. Box plots show gene expression (reads per kilobase per million (RPKM)) values as measured by RNA-seq¹⁸ for assigned target genes (blue) and non-regulated host genes (red, left) or skipped genes (red, right). Dark grey, unrelated neighbouring genes (control); light grey, all *D. melanogaster* genes. *** $P = 10^{-8}$, ** $P = 0.059$, * $P = 0.081$, ~ $P > 0.1$. Wilcoxon rank-sum test.

fraction of classes for which predictions were successful increased with the number of enhancers per class (Supplementary Table 5) but appeared to be independent of pattern complexity. This suggests that our understanding of regulatory sequences will benefit from the ongoing functional characterization of enhancers^{4–7}.

Different transcription factor motifs were strongly differentially distributed between the enhancer classes (Fig. 4a and Extended Data Fig. 9c). For example, early embryonic enhancers were enriched in motifs of the transcription factor Zelda, an important activator of embryonic gene expression³⁰. Similarly, Twist (Twi) motifs were enriched in early mesodermal enhancers, Myocyte enhancing factor 2 (Mef2) motifs in late

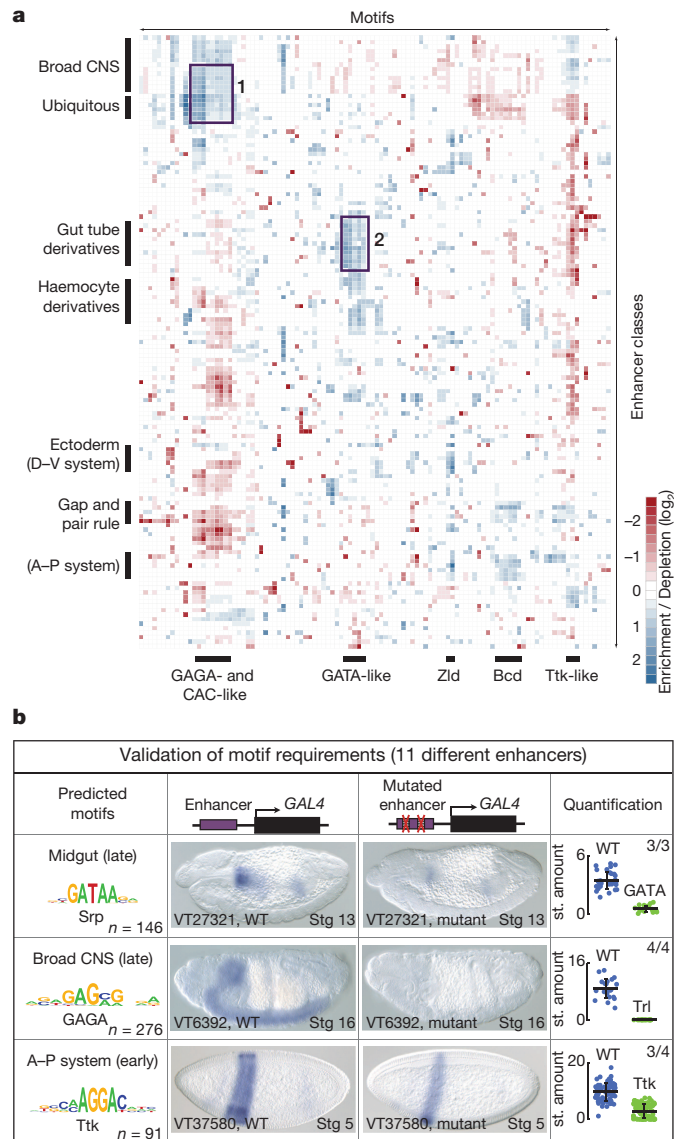


Figure 4 | Prediction and validation of *cis*-regulatory motif requirements for tissue-specific enhancer activities. **a**, Global *cis*-regulatory map of transcription factor motif enrichments in sequences of enhancers active in different tissues/cell types. Highlighted are Trl (GAGA) and CAC(N)_NCAC-like motifs enriched in CNS and ubiquitous enhancers (1) and GATA-like motifs enriched in midgut enhancers (2; see Extended Data Fig. 9c for the entire map). D–V, dorso–ventral. Zld, Zelda. **b**, Experimental validation of predicted *cis*-regulatory motif requirements. Shown are the most discriminative motifs (left), representative enhancers active in the midgut (stages 13–15), broad CNS (stages 15–16) and A–P system (stages 4–6) and their motif mutant variants (middle), and a quantification of the staining (st) intensities (right; all $P \leq 7 \times 10^{-10}$, Kolmogorov–Smirnov; see Extended Data Figs 9a, b and 10 for details and eight additional enhancers). WT, wild type.

somatic muscle enhancers, and Pannier (Pnr) and Tinman (Tin) motifs in dorsal vessel enhancers, consistent with the established roles of these transcription factors⁸ (Fig. 4a and Extended Data Fig. 9c). To test whether predicted motifs are required for enhancer activity, we selected three midgut, four CNS and four anterior–posterior (A–P) enhancers (11 enhancers total), for which the successful predictions depended on GATA-like, Trithorax (Trl, also known as GAGA)-like, and Tramtrack (Ttk)-like motifs, respectively (Fig. 4a and Extended Data Fig. 9b, c). For each, we created reporter flies with an enhancer variant in which we disrupted the respective motifs by point mutations and compared the activity of the mutant and wild-type enhancers, both manually and by computational image analysis (Fig. 4b and Extended Data Fig. 10). In 10 out of 11 cases, the mutated enhancers were not active or had strongly reduced activity, validating the functional importance of the respective motifs.

Taken together this work complements efforts that study chromatin properties^{2,19,20} or characterize enhancers at defined stages and in selected tissues^{5–7,15,16}. Our results confirm and generalize principles and models from smaller scale studies (reviewed in refs 1, 9, 12) and suggest a high density of developmental enhancers in the *Drosophila* genome with an estimated total of ~41,000 enhancers or four enhancers per expressed protein-coding gene on average during embryogenesis alone. In addition, considering that enhancers that are exclusively active in larvae, pupae or the adult fly^{5–7,15,16} (Supplementary Information section 6), we estimate between at least 50,000 to 100,000 developmental enhancers in the 170-megabase *D. melanogaster* genome. Even though the genome sequence properties (for example, repeat content and gene density) differ, this suggests that the 3-gigabase human genome could contain up to several million enhancers. In summary, the functional characterization of enhancers during the entire *Drosophila* embryogenesis adds a new level of functional annotation to the well-studied fly genome and elucidates global principles of *cis*-regulatory genome organization in animals, the importance of which for development, physiology, evolution and disease is becoming increasingly evident.

METHODS SUMMARY

We assessed enhancer activities of 7,705 genomic fragments of about 2 kb in embryos of transgenic *GAL4*-reporter (VT) fly strains obtained from the VDRC (<http://stockcenter.vdrc.at/>) by *in situ* hybridization. Embryos of each VT strain were manually annotated with a controlled vocabulary and positive strains were imaged. Motif analyses and support vector machine (SVM) predictions were performed as described in ref. 29. All fragment coordinates and annotations are in Supplementary Table 1 and at <http://enhancers.starklab.org/>.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January; accepted 17 April 2014.

Published online 1 June 2014.

- Levine, M. Transcriptional enhancers in animal development and evolution. *Curr. Biol.* **20**, R754–R763 (2010).
- The modENCODE Consortium. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **489**, 57–74 (2010).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
- Jenett, A. *et al.* A *GAL4*-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).
- Manning, L. *et al.* A resource for manipulating gene expression and analyzing *cis*-regulatory modules in the *Drosophila* CNS. *Cell Rep.* **2**, 1002–1013 (2012).
- Jory, A. *et al.* A survey of 6,300 genomic fragments for *cis*-regulatory activity in the imaginal discs of *Drosophila melanogaster*. *Cell Rep.* **2**, 1014–1024 (2012).
- Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. M. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462**, 65–70 (2009).
- Yáñez-Cuna, J. O., Kvon, E. Z. & Stark, A. Deciphering the transcriptional *cis*-regulatory code. *Trends Genet.* **29**, 11–22 (2013).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).

11. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **448**, 116–120 (2012).
12. Zeitlinger, J. & Stark, A. Developmental gene regulation in the era of genomics. *Dev. Biol.* **339**, 230–239 (2010).
13. Rubin, G. M. & Lewis, E. B. A brief history of *Drosophila*'s contributions to genome research. *Science* **287**, 2216–2218 (2000).
14. Tomancak, P. *et al.* Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, RESEARCH0088 (2002).
15. Gallo, S. M. *et al.* REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.* **39**, D118–D123 (2011).
16. Pfeiffer, B. D. *et al.* Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105**, 9715–9720 (2008).
17. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, 1997).
18. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
19. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genet.* **44**, 148–156 (2012).
20. Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12**, R43 (2011).
21. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
22. Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6**, e1000814 (2010).
23. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
24. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314 (2008).
25. Calhoun, V. C., Stathopoulos, A. & Levine, M. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila* Antennapedia complex. *Proc. Natl Acad. Sci. USA* **99**, 9243–9247 (2002).
26. Jack, J., Dorsett, D., Delotto, Y. & Liu, S. Expression of the cut locus in the *Drosophila* wing margin is required for cell type specification and is regulated by a distant enhancer. *Development* **113**, 735–747 (1991).
27. Maeda, R. K. & Karch, F. Gene expression in time and space: additive vs hierarchical organization of *cis*-regulatory regions. *Curr. Opin. Genet. Dev.* **21**, 187–193 (2011).
28. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the *Fgf8* regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
29. Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D. & Stark, A. Uncovering *cis*-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res.* **22**, 2018–2030 (2012).
30. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**, 400–403 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Dickson laboratory VT project for cloning the candidate regions and generating transgenic flies and the VDRC (<http://stockcenter.vdrc.at>) for their maintenance and distribution. We are grateful to the IMP/IMBA Scientific Services, in particular BioOptics, Genomics, and IT for help and to C. H. Lampert (IST Austria) for advice. We thank M. Levine (UC Berkeley) and V. Hartenstein (UCLA) for their permission to reproduce figures. The Stark group is supported by a European Research Council (ERC) Starting Grant from the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 242922 awarded to A.S. and by the Austrian Science Fund (FWF, F4303-B09). Generation of transgenic lines was supported in part by an ERC Advanced Investigator Grant to B.J.D. Basic research at the IMP is supported by Boehringer Ingelheim GmbH.

Author Contributions E.Z.K. and A.S. conceived the study and wrote the paper. E.Z.K., G.S., M.P. and K.S. performed the screen, E.Z.K. annotated all patterns and performed validation experiments, E.Z.K. and T.K. performed the imaging, T.K. and G.S. analysed the image data, T.K. developed <http://enhancers.starklab.org>, J.O.Y.-C. performed all sequence analyses, B.J.D. provided transgenic flies prior to publication, E.Z.K., T.K., G.S., J.O.Y.-C. and A.S. analysed data. A.S. supervised the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. (stark@starklab.org).

Enhancer loops appear stable during development and are associated with paused polymerase

Yad Ghavi-Helm¹, Felix A. Klein^{1*}, Tibor Pakozdi^{1*}, Lucia Ciglar¹, Daan Noordermeer², Wolfgang Huber¹ & Eileen E. M. Furlong¹

Developmental enhancers initiate transcription and are fundamental to our understanding of developmental networks, evolution and disease. Despite their importance, the properties governing enhancer–promoter interactions and their dynamics during embryogenesis remain unclear. At the β -globin locus, enhancer–promoter interactions appear dynamic and cell-type specific^{1,2}, whereas at the *HoxD* locus they are stable and ubiquitous, being present in tissues where the target genes are not expressed^{3,4}. The extent to which preformed enhancer–promoter conformations exist at other, more typical, loci and how transcription is eventually triggered is unclear. Here we generated a high-resolution map of enhancer three-dimensional contacts during *Drosophila* embryogenesis, covering two developmental stages and tissue contexts, at unprecedented resolution. Although local regulatory interactions are common, long-range interactions are highly prevalent within the compact *Drosophila* genome. Each enhancer contacts multiple enhancers, and promoters with similar expression, suggesting a role in their co-regulation. Notably, most interactions appear unchanged between tissue context and across development, arising before gene activation, and are frequently associated with paused RNA polymerase. Our results indicate that the general topology governing enhancer contacts is conserved from flies to humans and suggest that transcription initiates from preformed enhancer–promoter loops through release of paused polymerase.

Drosophila embryogenesis proceeds very rapidly, taking 18 h from egg lay to completion. Underlying this dynamic developmental program are marked changes in transcription, which are in turn regulated by characterized changes in enhancer activity. However, the role and extent of dynamic enhancer looping during this process remains unknown. To address this, we performed 4C-seq (chromosome conformation capture sequencing) experiments⁵ anchored on 103 distal or promoter-proximal developmental enhancers (referred to as ‘viewpoints’; Extended Data Fig. 1a), and constructed absolute and differential interaction maps for each, varying two important parameters: (1) developmental time, using embryos at two different stages, early in development when cells are multipotent (3–4 h after egg lay; stages 6–7), and mid-embryogenesis during cell-fate specification (6–8 h; stages 10–11); and (2) tissue context, comparing enhancer interactions in mesodermal cells versus whole embryo. To perform cell-type-specific 4C-seq in embryos, we established a modified version of BiTS-ChIP (batch isolation of tissue-specific chromatin for immunoprecipitation)⁶. Nuclei from covalently crosslinked transgenic embryos, expressing a nuclear-tagged protein only in mesodermal cells, were isolated by fluorescence-activated cell sorting (FACS; (>98% purity) and used for 4C-seq on 92 enhancers at 6–8 h and a subset of 14 enhancers at 3–4 h. The same 92 enhancers, and 11 additional regions, were also used as viewpoints in whole embryos at both time points (Extended Data Fig. 1b and Supplementary Table 1). The enhancers were selected based on dynamic changes in mesodermal transcription factor occupancy between these developmental stages^{7,8} and the expression of the closest gene⁹. We thereby primed this study to detect dynamic three-dimensional (3D) interactions, focusing on developmental

stages during which the embryo undergoes marked morphological and transcriptional changes.

All 4C-seq experiments had the expected signal distribution⁵, with high concordance between replicates (median Spearman correlation 0.93). To assess data quality further, we examined ten known enhancer–promoter pairs (of the *ap*, *Abd-b*, *E2f*, *pdm2*, *Con*, *eya*, *stumps*, *Mef2*, *sli* and *slp1* genes), and in all cases recovered the expected interactions (Fig. 1 and Extended Data Fig. 1c–l). For example, using an enhancer of the *apterous* (*ap*) gene¹⁰, we detect the expected interaction with the *ap* promoter, 17 kilobases (kb) away (Fig. 1), illustrating the high quality and resolution of the data.

In chromosome conformation capture assays, interaction frequencies decrease with genomic distance between regions⁵. To adjust for this, we modelled the 4C signal decay as a function of distance using a monotonously decreasing smooth function¹¹ (Extended Data Fig. 1b). Subtracting this trend, the residual interaction signal was converted to z-scores and interacting regions defined by merging neighbouring high-scoring fragments within 1 kb. Using this stringent approach, 4,247 high-confidence interactions were identified across all viewpoints and conditions, representing 1,036 unique interacting regions (Supplementary Table 2).

Each enhancer (viewpoint) interacted with, on average, ten distinct genomic regions (Extended Data Fig. 2a), less than half (41%) of which were annotated enhancers or promoters. Distal enhancers had a higher than expected interaction frequency with other enhancers (Extended Data Fig. 2b, $P = 2.4 \times 10^{-3}$). Similarly, promoter-proximal elements

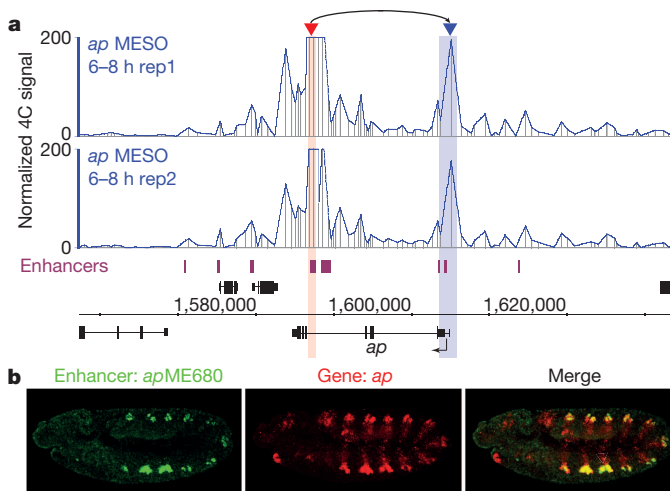


Figure 1 | A high-resolution view of enhancer interactions during *Drosophila* embryogenesis. **a**, 4C interaction map (viewpoint, red arrowhead) at the *ap* locus. The expected interaction with the promoter (blue arrowhead) of *ap* is observed. Known enhancers are indicated. **b**, Expression (*in situ* hybridization) of the *ap* gene (red) and expression driven by its interacting enhancer (GFP, green) at stage 11. MESO, mesoderm.

¹European Molecular Biology Laboratory, Genome Biology Unit, D-69117 Heidelberg, Germany. ²Swiss Federal Institute of Technology, School of Life Sciences, CH-1015 Lausanne, Switzerland.

*These authors contributed equally to this work.

had extensive interactions with distal active promoters, 98% of which are >10 kb away (Extended Data Fig. 2b, c, $P = 6 \times 10^{-4}$). Enhancer-promoter interactions, although not significantly enriched, involve active promoters, with high enrichment for H3K27ac and H3K4me3, and active enhancers, defined by H3K27ac, RNA Pol II and H3K79me3 (ref. 6) (Extended Data Fig. 2d, e). In contrast, contacts at inactive promoters are significantly depleted (Extended Data Fig. 2b). These results are similar to recent findings in human cells^{12,13} and the mouse β -globin locus^{1,2}, indicating similarities in 3D regulatory principles from flies to humans.

The extent of 3D connectivity is surprising given the relative simplicity of the *Drosophila* genome. On average, each promoter-proximal element interacted with four distal promoters and two annotated enhancers, whereas each distal enhancer interacted with two promoters and three other enhancers. These numbers are probably underestimates, as 60% of interactions involved intragenic or intergenic fragments containing no annotated *cis*-regulatory elements. Despite this, the level of connectivity is similar to that recently observed in humans, where active promoters contacted on average 4.75 enhancers and 25% of enhancers interacted with two or more promoters¹³. The multi-component contacts that we observe for *Drosophila* enhancers indicate topologically complex structures and suggest that, despite its non-coding genome being an order of magnitude smaller than humans, *Drosophila* may require a similar 3D spatial organization to ensure functionality.

Insulators, and associated proteins, are thought to have a major role in shaping nuclear architecture by anchoring enhancer-promoter interactions or by acting as boundary elements between topologically associated domains (TADs)^{14–16}. Occupancy data from 0 to 12 h *Drosophila* embryos¹⁷ revealed a 50% overlap of interacting regions with occupancy of one or more insulator protein. Insulator-bound interactions are enriched in enhancer elements, suggesting that insulators may have a role in promoting enhancer-enhancer interactions (Extended Data Fig. 3a–d). In contrast to mammalian cells¹⁶, we observed no association between insulator occupancy and the genomic distance spanned by chromatin loops, although there was a modest increase in average interaction strength (Extended Data Fig. 3e, f). Conversely, 50% of interacting regions are not bound by any of the six *Drosophila* insulator proteins (Extended Data Fig. 3a, g), suggesting that these 3D contacts are formed in an insulator-independent manner, or are being facilitated by neighbouring interacting regions.

If enhancer 3D contacts are involved in transcriptional regulation, then genes linked by interactions with a common enhancer should share spatio-temporal expression, as recently reported^{18–20}. For the four loci examined—*pdm2* (Extended Data Fig. 4a, b), *engrailed* (*en*; Extended Data Fig. 4c, d), *unc-5* (Extended Data Fig. 5c, d) and *charybde* (Fig. 2c, d, described below)—this is indeed the case. For example, the *pdm2* CE8012 enhancer interacts with both the *pdm2* and *nubbin* (*nub*), also known as *pdm1* promoters, located 2.5 and 47 kb away, respectively. Both genes, producing structurally related proteins, are co-expressed in the ectoderm, overlapping the activity of the *pdm2* enhancer.

Although there are examples of long-range interactions in *Drosophila*, often involving Polycomb response elements (PREs)^{15,21,22} and insulator elements²¹, the vast majority of characterized enhancers are within 10 kb of their target gene, with few known to act over 50 kb (Fig. 2a and Supplementary Table 3). However, as investigators historically tested regions close to the gene of interest, characterized *Drosophila* enhancers are generally close to the gene they regulate. In contrast, although 4C cannot assess the full extent of short-range interactions (Extended Data Fig. 5a, b), it provides an unbiased systematic measurement of the distance of enhancer interactions, far beyond 10 kb.

The distance distribution of all significant interactions reveals extensive long-range interactions within the ~180 megabase (Mb) *Drosophila* genome; 73% span >50 kb, with the median interaction-viewpoint distance being 110 kb (Fig. 2a, b). Two striking examples of long-range interactions are the *unc-5* and *charybde* loci. The *unc-5* promoter interacts with multiple regions, including a weak but significant interaction

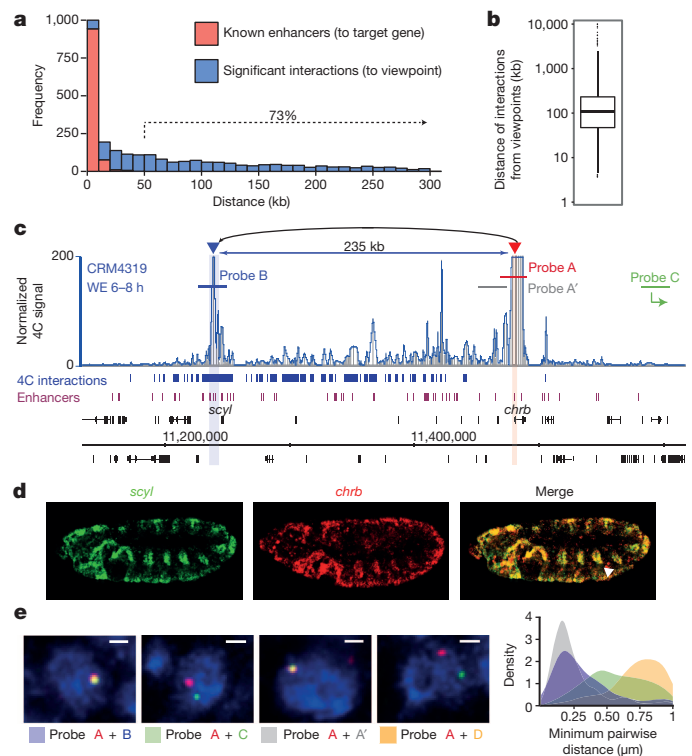


Figure 2 | Long-range interactions are widespread in the *Drosophila* genome. **a**, **b**, Distance distribution of all known *Drosophila* enhancers (red; literature based) and identified significant interactions (blue, $n = 1,983$) to their respective target gene or viewpoint, shown as a histogram (**a**) and box plot (**b**). **c**, 4C interaction map (viewpoint, red arrowhead) around the *scyl* and *chrB* loci. WE, whole embryo. Interaction with the *scyl* gene is highlighted (blue arrowhead). Significant 4C interactions, known enhancers and DNA FISH probes are indicated. **d**, Expression (*in situ* hybridization) of *chrB* (red) and *scyl* (green) at stage 11. **e**, DNA FISH images of representative nuclei, merging DAPI (blue), probe A (red) and probe B, C, A' or D (green) channels. Scale bar, 1 μ m. Density plot indicates measured distances between probe A and probe A' (grey), D (beige) B (blue) or C (green).

with the promoter of *slit* (*slit*), at a distance of >500 kb (Extended Data Fig. 5c, d). These genes produce structurally unrelated proteins that are co-expressed in the heart, and are essential for heart formation.

A promoter-proximal element near the *charybde* (*chrB*) promoter has a strong interaction with the promoter of the *scylla* (*scyl*) gene, almost 250 kb away (Fig. 2c). Both genes are closely related in sequence and co-expressed throughout embryogenesis (Fig. 2d)²³. These long-range interactions were confirmed by reciprocal 4C, using either the promoter of *chrB* or *scyl*, or an interacting putative enhancer as viewpoint (Extended Data Fig. 5e). We further verified this interaction using DNA fluorescence *in situ* hybridization (FISH) in embryos (Fig. 2e). As a control, we assessed the distance between the *chrB* promoter (probe A) and an overlapping probe A' or a region on another chromosome (probe D), to determine the distances between regions very close or far away, respectively. Comparing the distance between the *chrB* and *scyl* promoters (probes A and B, Fig. 2c) showed a high, statistically significant co-localization (Fig. 2e; 37% co-localization; $P < 10^{-18}$; Extended Data Fig. 5f), in contrast to the distance between the *chrB* promoter and a non-interacting region with equal genomic distance (probes A and C; 5% co-localization).

The reciprocal 4C revealed several intervening interactions that are consistently associated with loops to both the *scyl* and *chrB* promoter. We examined the activity of two of these in transgenic embryos. Both interacting regions can function as enhancers *in vivo*, recapitulating *chrB* expression in the visceral mesoderm (enhancer 1) and nervous system (enhancer 2) (Extended Data Fig. 5e, g).

When considering a 1-Mb scale around this region, the 4C interaction signal drops to almost zero just after the promoters of both genes (Extended Data Fig. 6a). This 'contained block' of interactions is reminiscent of TADs¹⁴, although the boundaries don't exactly match TADs defined at late stages of embryogenesis¹⁵, which may reflect differences in the developmental stages used. However, the boundaries do overlap a block of conserved microsynteny between drosophilids²⁴ spanning ~50 million years of evolution (Extended Data Fig. 6a), suggesting a functional explanation underlying the maintained synteny. Expanding this analysis across all viewpoints, ~60% of interactions are located within the same TAD and the same microsyntenic domain as the viewpoint (Extended Data Fig. 6b, c). In the case of the *chrB* and *scyl* genes, this constraint may act to maintain a regulatory association between a large array of enhancers, facilitating their interaction with both genes' promoters.

These examples, and the other 555 unique interactions >100 kb, provide strong evidence that long-range interactions are widely used within the *Drosophila* genome, potentially markedly increasing the regulatory repertoire of each gene.

As enhancer–promoter looping can trigger gene expression²⁵, it follows that enhancer contacts should reflect the dynamics of transcriptional changes during development and therefore be temporally associated with gene expression. To assess this, we directly compared looping interactions between the two different time points and tissue contexts. Given the non-discrete nature of chromatin contacts, we used the quantitative 4C-seq signal to identify differential interactions based on a Gamma-Poisson model and defined them as having >2-fold change and false discovery rate $\leq 10\%$.

Despite the marked differences in development and enhancer activity between these conditions, we found surprisingly few changes in chromatin interaction frequencies, with ~6% of interacting fragments showing significant changes between conditions (Extended Data Fig. 7; Fig. 3a

and Extended Data Fig. 8a, red dots). Of these, 87 interactions were significantly reduced during mid-embryogenesis (6–8 h) compared to the early time point (3–4 h), and 90 interactions significantly increased. Similarly, 105 interactions had a higher frequency in mesodermal cells, compared to the whole embryo, and 34 interactions were lower.

For example, a promoter-proximal viewpoint in the vicinity of the *Antp* promoter identified many interactions, two of which are significantly decreased at 6–8 h, although the expression of the *Antp* gene itself increases (Extended Data Fig. 8b). For one region, the reduction in 4C interaction at 6–8 h corresponds to a loss in a H3K4me3 peak from 3–4 h to 6–8 h (asterisk), suggesting that this 3D contact is associated with the transient expression of an unannotated transcript. We examined the activity of the other interacting peak in transgenic embryos and showed that it acts as an enhancer, driving specific expression in the nervous system overlapping the *Antp* gene at 6–8 h (Extended Data Fig. 8c). Along with the two enhancers discovered at the *chrB* locus, this demonstrates the value of 3D interactions to identify new enhancer elements, even for well-characterized loci like *Antp*.

A viewpoint in the vicinity of the *Abd-B* promoter interacted with a number of regions spanning the bithorax locus, three of which correspond to previously characterized *Abd-B* enhancers; iab-5 (ref. 26), iab-7 and iab-8 (refs 26, 27) (Fig. 3b, c). The iab-7 and iab-8 enhancers are active in early embryogenesis, and have much reduced or no activity at the later time point^{26,27}. Notably, although the loop to those two enhancers is strong at the early time point, it becomes significantly reduced later in development, when both enhancers' activities are reduced. Conversely, the iab-5 enhancer contacts the promoter at a much higher frequency later in development, at the stage when the enhancer is most active^{26,27}. This locus therefore exhibits dynamic 3D promoter–enhancer contacts that reflect the transient activity of three developmental enhancers. It is interesting to note that in all loci examined, the dynamic contacts of specific elements are neighboured by stable contacts, as seen

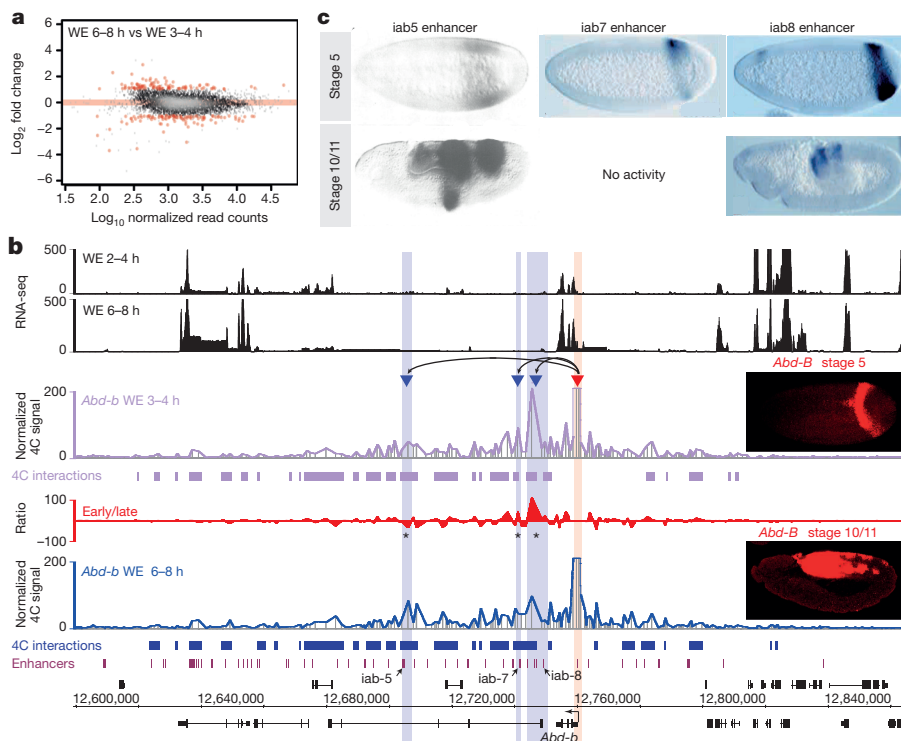


Figure 3 | Specific loci display localized differential interactions. **a**, MA plot of interaction signal between whole embryo 6–8 h and whole embryo 3–4 h (significant differential interactions, red dots). WE, whole embryo. **b**, 4C interaction map at the *Abd-b* locus. Top to bottom: RNA-seq signal (reads per kilobase per million mapped reads (RPKM), black) in whole embryo at 2–4 h and 6–8 h (ref. 9), 4C interaction map (viewpoint, red arrowhead) in whole

embryo at 3–4 h (mauve) and 6–8 h (blue), and differential 4C signal (red) with significant differential 4C interactions (asterisk). Insets show the expression (*in situ* hybridization) of *Abd-B* (red) at stages 5 (2–4 h) and 11 (6–8 h). **c**, Expression (*in situ* hybridization) driven by iab5 (ref. 26), iab7 and iab8 (ref. 27) enhancers at stages 5 and 10–11. Embryo images in panel c reproduced with permission from: ref. 27, *Development*; ref. 26, Nature Publishing Group.

in the *Antp* and *Abd-B* loci. Dynamic changes, therefore, appear to operate in the context of larger, more-stable 3D landscapes.

Ninety-four per cent of enhancer interactions showed no evidence of dynamic changes across time and tissue context, which is remarkable given the marked developmental transitions during these stages (Fig. 3a, Extended Data Fig. 8a and Supplementary Table 4). To investigate this further, we examined enhancer–promoter interactions of genes switching their expression state between time points or tissue contexts. The *ap* gene, for example, is not expressed at 2–4 h but is highly expressed during mid-embryogenesis (6–8 h) (Fig. 4a). Despite the absence of expression, the interaction between the *apME680* enhancer and the *ap* promoter is already present at 3–4 h, several hours before the gene's activation (Fig. 4a). To examine this more globally, we selected differentially expressed genes, going either from on-to-off or off-to-on (Extended Data Fig. 9). Even for these dynamically expressed genes, there was no correlation with changes in their promoter–enhancer contacts (Fig. 4b). We observe similar 'stable' interactions between tissue contexts. Genes predominantly expressed in the neuroectoderm at 6–8 h, for example, have interactions at the same locations in whole embryos and purified mesodermal nuclei at 6–8 h, despite the fact that they are not expressed in the mesoderm at this stage (Extended Data Fig. 8d–g).

Pre-existing loops were recently observed in human and mouse cells, and suggested to prime a locus for transcriptional activation^{4,13}. However, why they are formed and how transcription is eventually triggered remains unclear. To investigate this, we focused on the subset of genes that have both off-to-on expression and no evidence for differential interactions (20 genes; differentially expressed with stable loops (DS) genes; Supplementary Table 5 and Extended Data Fig. 9). Despite changes in their overall expression, DS genes have similar levels of RNA polymerase II (Pol II) promoter occupancy at both time points (Fig. 4c).

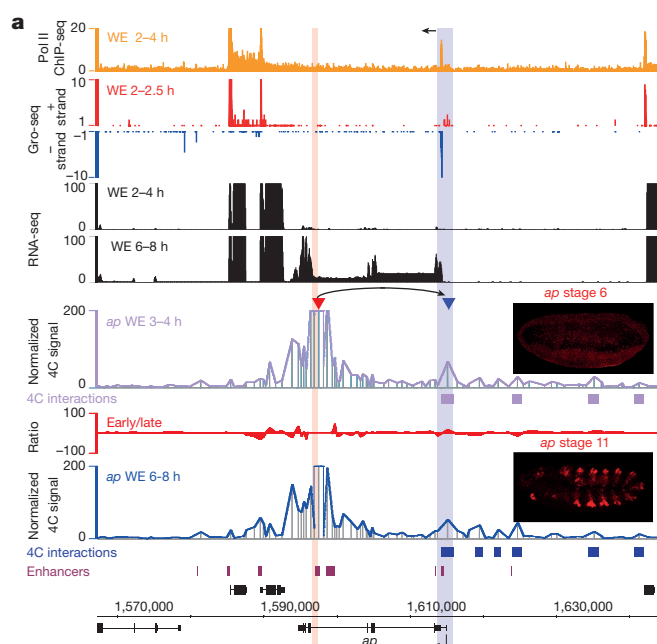
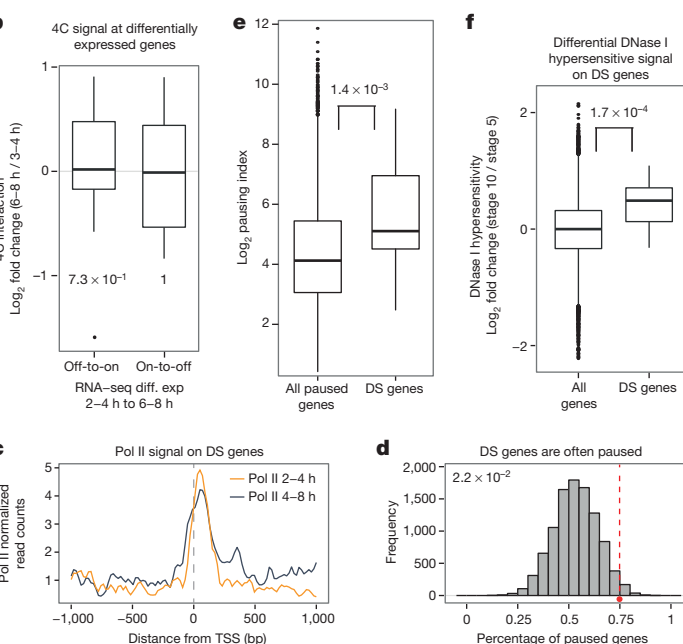


Figure 4 | Interactions are stable across developmental time and associated with paused polymerase. **a**, 4C interaction map at the *ap* locus. Top to bottom: Pol II signal (reads per genomic content (RPGC)) in whole embryo at 2–4 h (orange), GRO-seq signal in whole embryo at 2–2.5 h (plus strand, red; minus strand, blue)²⁸, RNA-seq signal (RPKM) in whole embryo at 2–4 h and 6–8 h (black)⁹, 4C interaction map (viewpoint, red arrowhead) in whole embryo at 3–4 h (mauve) and 6–8 h (blue), and differential 4C signal (red). WE, whole embryo. Significant 4C interactions and known enhancers are indicated. Insets show the expression (*in situ* hybridization) of *ap* at stages 6 (3–4 h) and 11 (6–8 h). **b**, Differentially expressed genes, going from off-to-on ($n = 21$) or on-to-off ($n = 8$), have no significant differences in the frequency of 4C interactions at their promoter (two-sided Wilcoxon test). **c**, Pol II signal (RPGC) is enriched at the promoter of differential genes with stable

The presence of promoter-bound Pol II in the absence of full-length transcription is indicative of Pol II pausing. Using global run-on sequencing (GRO-seq) data²⁸ to define a stringent set of paused genes, we observed that most (75%) DS genes are paused (15 of 20 DS genes; Fig. 4d and Extended Data Fig. 9b, d), and have a significantly higher pausing index (Fig. 4e). This percentage is significantly higher than expected by chance when sampling over all off-to-on genes (Fig. 4d), and is robust to using a strict (Fig. 4d) or more relaxed (Extended Data Fig. 9e) definition of Pol II pausing²⁸. This association is very evident when examining specific loci (Fig. 4a and Extended Data Fig. 10), showing Pol II occupancy, short abortive transcripts, and loop formation before the gene's expression. Taken together, these results indicate that 'stable' chromatin loops are associated with the presence of paused Pol II at the promoter.

To understand how transcription is ultimately activated, we examined changes in DNase I hypersensitivity²⁹ at the promoter of DS genes. DNase I hypersensitivity is significantly increased at interacting promoters at the stages when the gene is expressed (Fig. 4f), suggesting that the recruitment of additional transcription factor(s) later in development might act as the trigger for transcriptional activation.

In summary, our data reveals extensive long-range interactions in an organism with a relatively compact genome, including pairs of co-regulated genes contacting common enhancers often at distances greater than 200 kb. Comparing enhancer contacts in different contexts revealed that chromatin interactions are very similar across developmental time points and tissue contexts. Enhancers therefore do not appear to undergo long-range looping *de novo* at the time of gene expression, but are rather already in close proximity to the promoter they will regulate. Within this 3D topology, highly dynamic and transient contacts would not be visible when averaging over millions of nuclei. As transcription factor



interactions (DS genes) at 2–4 h, even though the genes are not expressed. **d**, Expected distribution of paused genes (using top 50% of paused genes²⁸) by sampling 10,000 times differential off-to-on genes (red dotted line, observed percentage of paused DS genes). P value is indicated. **e**, Log₂ pausing index in whole embryo at 2–2.5 h (ref. 28) of DS genes ($n = 19$) is significantly different from all paused genes²⁸ ($n = 7,734$; two-sided Wilcoxon test). **f**, Box plot showing differential DNase I hypersensitivity (log₂ fold change stage 10/stage 5) at the promoter of DS genes ($n = 18$) and all *Drosophila* mRNA genes ($n = 10,409$, see Methods). P value from a two-sided Wilcoxon test. Boxes depict the interquartile range (IQR) with the median as a horizontal thick line. Upper and lower whiskers extend to 1.5 times the IQR, and points represent outliers.

binding is sufficient to force loop formation²⁵, our results suggest a model where through transcription factor–enhancer occupancy, an enhancer loops towards the promoter and polymerase is recruited, but paused in the majority of cases. The subsequent recruitment of transcription factor(s) or additional enhancers at preformed 3D hubs most likely triggers activation by releasing Pol II pausing. Such preformed topologies could thereby promote rapid activation of transcription^{3,4,13}. At the same time, as paused promoters can exert enhancer-blocking activity³⁰, the presence of paused polymerase within these 3D landscapes could safeguard against premature transcriptional activation, but yet keep the system poised for activation.

METHODS SUMMARY

Staged *Drosophila* embryos were collected at 3–4 h or 6–8 h after egg lay and fixed in 1.8% formaldehyde for 15 min at room temperature. Thirty million nuclei were used for each 4C template preparation, enough for on average ten viewpoints. Libraries were amplified from 320 ng of 4C template (primer sequence in Supplementary Table 6), and 100 multiplexed libraries were sequenced over on average five HiSeq2000 lanes, using 100-base-pair (bp) single-end reads. Two independent biological replicates were analysed for each condition.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 December 2013; accepted 29 April 2014.

Published online 2 July 2014.

1. Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
2. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C). *Nature Genet.* **38**, 1348–1354 (2006).
3. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
4. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
5. Van de Werken, H. J. G. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
6. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genet.* **44**, 148–156 (2012).
7. Junion, G. *et al.* A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**, 473–486 (2012).
8. Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. M. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65–70 (2009).
9. Graveley, B. R. *et al.* The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479 (2011).
10. Capovilla, M., Kambris, Z. & Botas, J. Direct regulation of the muscle-identity gene *apterous* by a Hox protein in the somatic mesoderm. *Development* **128**, 1221–1230 (2001).
11. Tolhuis, B. *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* **7**, e1001343 (2011).
12. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
13. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
14. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
15. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
16. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
17. Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6**, e1000814 (2010).
18. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42**, 53–61 (2010).
19. Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M. S. & Mhlanga, M. M. Chromosomal contact permits transcription between coregulated genes. *Cell* **155**, 606–620 (2013).
20. Zhang, Y. *et al.* Chromatin connectivity maps reveal dynamic promoter–enhancer long-range associations. *Nature* **504**, 306–310 (2013).
21. Cléard, F., Moshkin, Y., Karch, F. & Maeda, R. K. Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nature Genet.* **38**, 931–935 (2006).
22. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. & Orlando, V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nature Cell Biol.* **9**, 1167–1174 (2007).
23. Scuderi, A., Simin, K., Kazuko, S. G., Metherall, J. E. & Letsou, A. *scylla* and *charybde*, homologues of the human apoptotic gene *RTP801*, are required for head involution in *Drosophila*. *Dev. Biol.* **291**, 110–122 (2006).
24. Engström, P. G., Sui, S. J. H., Drivenes, Ø., Becker, T. S. & Lenhard, B. Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908 (2007).
25. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
26. Busturia, A. & Bienz, M. Silencers in abdominal-B, a homeotic *Drosophila* gene. *EMBO J.* **12**, 1415–1425 (1993).
27. Zhou, J., Ashe, H., Burks, C. & Levine, M. Characterization of the transvection mediating region of the abdominal-B locus in *Drosophila*. *Development* **126**, 3057–3065 (1999).
28. Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T. & Ashe, H. L. Extensive polymerase pausing during *Drosophila* axis patterning enables high-level and pliable transcription. *Genes Dev.* **27**, 1146–1158 (2013).
29. Thomas, S. *et al.* Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* **12**, R43 (2011).
30. Chopra, V. S., Cande, J., Hong, J.-W. & Levine, M. Stalled Hox promoters as chromosomal boundaries. *Genes Dev.* **23**, 1505–1509 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was technically supported by the EMBL Genomics Core and FACS core facilities. We thank all members of the Furlong laboratory for discussions and comments, in particular D. Garfield, J. Reddington and I. Schor for important suggestions. Embryo images in Fig. 3c were used with permission from *Development*²⁷ and Nature Publishing Group²⁶ as these fly strains no longer exist. This work was supported by a DFG (FU 750) grant to E.E.M.F., an EMBO post-doctoral fellowship to Y.G.-H., and the EC FP7 project ‘Radiant’ grant to F.A.K. and W.H.

Author Contributions Y.G.-H. and E.E.M.F. designed the study, analysed the results and wrote the manuscript. Y.G.-H. performed 4C experiments, DNA *in situ* hybridization and imaging and performed data analysis. F.A.K., T.P. and W.H. developed and performed 4C-seq bioinformatics analysis. L.C. generated all transgenic strains and performed 4C-PCR reactions, RNA *in situ* hybridizations and imaging. D.N. was involved in 4C primer design. T.P. and F.A.K. contributed equally to the study. All authors discussed the results and commented on the manuscript.

Author Information All raw data, which consists of 2,587 demultiplexed files, have been submitted to the EBI European Nucleotide Archive and ArrayExpress databases, accession numbers ERP004524 and E-MTAB-2180, respectively. To enable the community to browse through the data, 4C-seq interaction data is available in a customized web browser at <http://furlonglab.embl.de/4CBrowser>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.E.M.F. (furlong@embl.de).

Visualizing the kinetic power stroke that drives proton-coupled zinc(II) transport

Sayan Gupta^{1†}, Jin Chai², Jie Cheng³, Rhijuta D'Mello¹, Mark R. Chance¹ & Dax Fu^{2,3}

The proton gradient is a principal energy source for respiration-dependent active transport, but the structural mechanisms of proton-coupled transport processes are poorly understood. YiiP is a proton-coupled zinc transporter found in the cytoplasmic membrane of *Escherichia coli*. Its transport site receives protons from water molecules that gain access to its hydrophobic environment and transduces the energy of an inward proton gradient to drive Zn(II) efflux^{1,2}. This membrane protein is a well-characterized member^{3–7} of the family of cation diffusion facilitators that occurs at all phylogenetic levels^{8–10}. Here we show, using X-ray-mediated hydroxyl radical labelling of YiiP and mass spectrometry, that Zn(II) binding triggers a highly localized, all-or-nothing change of water accessibility to the transport site and an adjacent hydrophobic gate. Millisecond time-resolved dynamics reveal a concerted and reciprocal pattern of accessibility changes along a transmembrane helix, suggesting a rigid-body helical re-orientation linked to Zn(II) binding that triggers the closing of the hydrophobic gate. The gated water access to the transport site enables a stationary proton gradient to facilitate the conversion of zinc-binding energy to the kinetic power stroke of a vectorial zinc transport. The kinetic details provide energetic insights into a proton-coupled active-transport reaction.

Mammalian homologues of YiiP are responsible for zinc sequestration into secretory vesicles, thus playing important roles in neurotransmission¹¹ and hormone secretion¹². Zinc efflux catalysed by YiiP is coupled with proton influx in a 1:1 zinc-for-proton exchange stoichiometry³. When protons are scarce at higher pH, zinc transport comes to a halt despite a large zinc concentration gradient⁷. Thus, the zinc-for-proton coupling is obligatory. Biochemical studies and X-ray structures of YiiP showed that zinc transport is mediated by a tetrahedral Zn(II)-binding site in the centre of the transmembrane domain (TMD)¹⁴. This intramembranous zinc-transport site adopts coordination geometry satisfied by three Asp and one His residues, but lacks any additional polar or charged residues in the Zn(II)-binding pocket. The absence of available pH titratable residues in the second coordination sphere necessitates water access to fulfil proton donor or acceptor functions to enable the obligatory zinc-for-proton exchange. However, the crystal structure of zinc-bound YiiP (zinc-YiiP) shows that water access to the transport site is blocked by hydrophobic residues that divide the zinc translocation pathway into an extracellular and intracellular cavity¹³. A protein conformational change is expected to open up a water portal within the hydrophobic seal. As water molecules gain access to the transport site in a transport reaction cycle, irradiating YiiP to a millisecond synchrotron X-ray pulse could render residues in contact with waters susceptible to hydroxyl-radical-mediated oxidative modification, thereby permitting the monitoring of residues motions in terms of water accessibility change^{14,15}. Radiolytic hydroxyl radicals under such experimental conditions are generated rapidly and isotropically both in bulk and activated bound waters with side-chain oxidation completed within milliseconds^{15–18}. By comparison, the macroscopic timescale for zinc transport is of the order of 200–500 ms^{3,7}. Thus, time-resolved hydroxyl radical 'footprinting' would have a sufficient time resolution to monitor proton translocation and associated protein conformational change.

Purified YiiP in detergent micelles was exposed to a focused synchrotron white beam, followed by a rapid mix with methionine-amide to quench secondary radical chain reactions (Extended Data Fig. 1a). The effective hydroxyl radical concentration was controlled in the micromolar range as indicated by an Alexa Fluor 488 dosimeter, and secondary radiation damage of YiiP was minimized by adjusting the X-ray irradiation to an optimal dose range^{14,16}. As a result, only negligible differences in size-exclusion high-performance liquid chromatography (HPLC) profiles were observed for the protein peaks before and after X-ray irradiation (Fig. 1a). The broad low molecular peak in zinc-YiiP (red trace) corresponded to the methionine-amide quencher added to the apo-YiiP sample after irradiation. The sites of oxidative modification were characterized by +14, +16 and +32 dalton (Da) oxygen-based mass adducts^{14,15,18}, which were detected by bottom-up liquid chromatography–mass spectrometry (LC–MS) of proteolytic fragments of the irradiated YiiP (Fig. 1b), and confirmed by tandem mass spectrometry (MS/MS) assignments (Fig. 1c). The overall mass spectrometric sequence coverage was 82% (Extended Data Fig. 2a), encompassing all residues located within the inter-cavity seal (Extended Data Fig. 2b). Increasing X-ray irradiation progressively increased the modified and reduced the unmodified populations, giving rise to a dose–response plot for each modified site (Fig. 1d and Extended Data Fig. 3). The initial phase of the dose–response plot followed a pseudo-first-order reaction, but occasional deviations from the exponential function were observed at increased irradiation times as a result of secondary modifications (Fig. 1d and Extended Data Fig. 3). Therefore, the slope of the initial phase was used to quantify the hydroxyl radical reactivity (Extended Data Table 1).

The rate of side-chain labelling is governed by intrinsic reactivity of the amino acid and water accessibility to the side chain^{14,15}. The ratio of the measured reactivity rates for the same residue from zinc-YiiP and apo-YiiP gave a ratiometric account of the water accessibility change independent of the intrinsic side-chain reactivity or sequence context. Among all the detectable sites of modification, two sites exhibited conspicuously large differences in reactivity in the presence and absence of zinc (Fig. 2a). One instance where Zn(II) binding reduced reactivity more than 1000-fold was observed for three consecutive residues, V48, D49 and I50, within the peptide LVDI of TM2 (Extended Data Table 1). In the crystal structure of zinc-YiiP (Protein Data Bank accession number 3H90), D49 binds Zn(II) in the transport site and is one helical turn away from a structural water that is immobilized via a hydrogen bond to S53 (3.1 Å to O_y) (Fig. 3a). Coordination of Zn(II) to the transport site may suppress productive radiolysis of this structural water, resulting in a negligible rate of VDI labelling in zinc-YiiP (Extended Data Table 1). In sharp contrast, the absence of a coordinated Zn(II) in apo-YiiP permitted an unusually fast radiolytic labelling at 163 s^{–1} (Extended Data Table 1). Such a high level of reactivity has been observed for radiolytic labelling by structural water molecules in the hydrophobic core of a G-protein-coupled receptor¹⁷.

A second very significant change was observed for a +14 Da modification of L152 in the peptide ADMLHY of TM5 (Fig. 2a). In apo-YiiP, the rate of +14 Da modification for L152 was 8.5 s^{–1} (Extended Data Table 1) while Zn(II) binding reduced the reactivity of L152 more than 100-fold, to a negligible level, illustrating that a Zn(II)-binding-induced conformational

¹Center for Synchrotron Biosciences and Center for Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, Ohio 44109, USA. ²Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA. ³Department of Physiology, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA. [†]Present address: Berkeley Center for Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

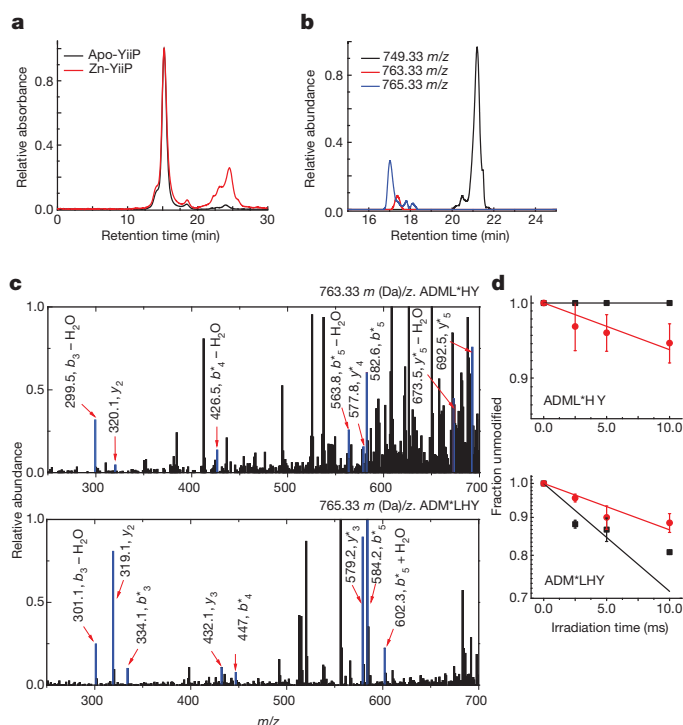


Figure 1 | Radiolytic labelling and mass spectrometric analysis. **a**, Size-exclusion HPLC chromatograms of apo-YiiP before irradiation and zinc-YiiP after irradiation. **b**, Examples for quantification of radiolytic labelling by LC-MS/MS; extracted ion-count chromatograms of singly protonated, unmodified (749.33 m/z, black), carbonylated (+14 Da mass shift, 763.33 m/z, red) and hydroxylated (+16 Da mass shift, 765.33 m/z, blue) peptide ADMLHY. **c**, Examples for identification of modified residues by MS/MS of the carbonylated and hydroxylated ADMLHY with peak assignments (red arrow and blue line) confirming L152 and M151 modification, respectively. **d**, Dose-responses showing reciprocal solvent accessibility changes at L152 and M151 sites in apo-YiiP (red) and zinc-YiiP (black). Solid lines, least-squares fits of the means of dose-dependent data; error bar, s.e.m. from four to six independent measurements.

change removed water access to the side chain of L152. This peptide contained another labelled residue, M151, whose +16 Da modified products could be isolated from those of L152 on the basis of the difference in the mass to charge ratio (m/z) (Fig. 1b, c). The same conformational change that reduced reactivity of L152 yielded a 2.5-fold increase in reactivity for the neighbouring M151 (Figs 1d and 2a). In the zinc-YiiP structure, L152 is fully buried and oriented towards the intracellular cavity as a part of the inter-cavity seal, consistent with the lack of radiolytic labelling (Fig. 3a).

L152 is located at the interface between a TM3–TM6 helix pair and a compact TM1–TM2–TM4–TM5 four-helix bundle (Fig. 3b). These two subdomains cross over to form two cavities located on either side of the membrane as indicated by arrows in Fig. 3b. The inter-domain packing wedges TM5 (coloured in red) at one corner of the four-helix bundle into the TM3–TM6 interface with L152 situated at the centre of the TM5→TM3–TM6 triple-helix joint (Fig. 3b). L152 interacts with I90 from TM3 and A194 from TM6 to form a tight knob-into-hole packing. The I90 equivalents have been identified as metal determinant residues in plant and yeast cation diffusion facilitator (CDF) homologues^{19,20}. One helical turn down towards the intracellular cavity is another layer of residue triad: A83 from TM3, A149 from TM5 and M197 from TM6, which define the innermost section of the intracellular cavity (Fig. 3b). Of note, the conformational changes of M197 echo those of L152, with a zinc-dependent reduction of solvent accessibility (Fig. 2a) except that the accessibility of M197 is not reduced to the background level in the zinc-bound state (Extended Data Fig. 3). This cluster of six residues forms a highly conserved TM5→TM3–TM6 packing core (Extended Data Fig. 4), with L152 serving as a principal hydrophobic barrier between the two

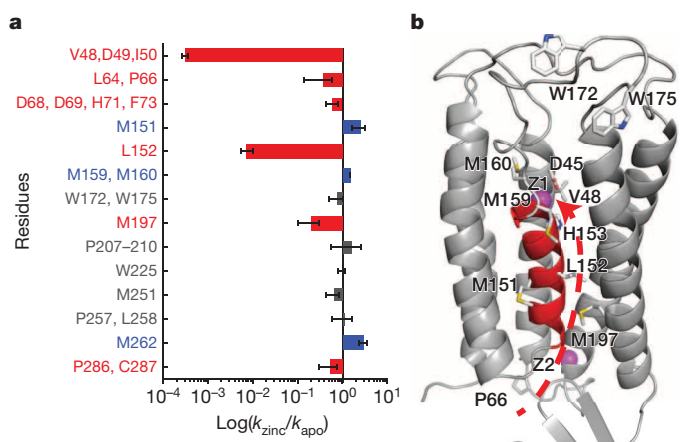


Figure 2 | Quantification of water accessibility changes. **a**, Water accessibility changes in response to Zn(II) binding measured by the ratio of labelling rates for residues with an increase (blue), decrease (red) or no change (grey) in water accessibility after a rapid Zn(II) exposure. The labelling rate for each site as indicated is summarized in Extended Data Table 1, and the error bar represents the standard error from four to six independent measurements. **b**, Residues with a partial water accessibility change in response to zinc binding. TM5 is coloured in red. Z1 and Z2 (magenta spheres) represent bound zinc ions. Arrow indicates a putative zinc-transport pathway from the cytoplasm through the L152 gate to the transport site.

cavities (Fig. 3c, d). The structural and functional importance of L152 was examined by a series of point mutations (Extended Data Fig. 5). All L152 mutants expressed well. However, substitutions of L152 with smaller (G, A), bulky aromatic (F) and charged residues (D, R) resulted in complete denaturation after the mutant proteins were solubilized by DDM, whereas conserved L152 substitutions with I and M residues were partly tolerated

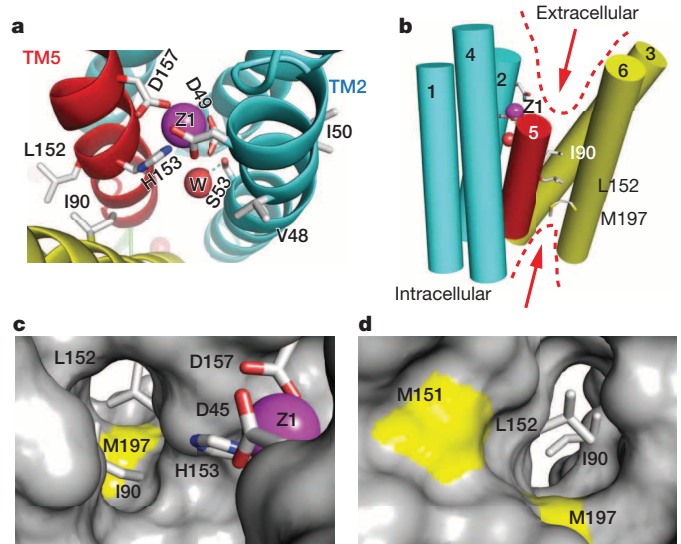


Figure 3 | L152 controls the opening of an inter-cavity water portal. **a**, A structural water molecule (W, red sphere) near the transport site occupied by a tetrahedrally coordinated Zn(II) (Z1, magenta sphere), viewed from the periplasm. Relevant residues are drawn in sticks and labelled accordingly. TM5 is coloured in red as indicated. **b**, Intracellular and extracellular cavity as outlined by dashed lines. **c**, L152 gate viewed from the extracellular cavity along the arrow as indicated in **b**. The side chains of L152, I90 and the coordination residues in the transport site (sticks) are excluded from the protein surface drawing. M197 is shown as a yellow patch at the cytoplasmic entrance to the inter-cavity portal. **d**, L152 gate viewed from the intracellular cavity along the arrow as indicated in **b**. M151 and M197 are visible as yellow patches on the protein surface.

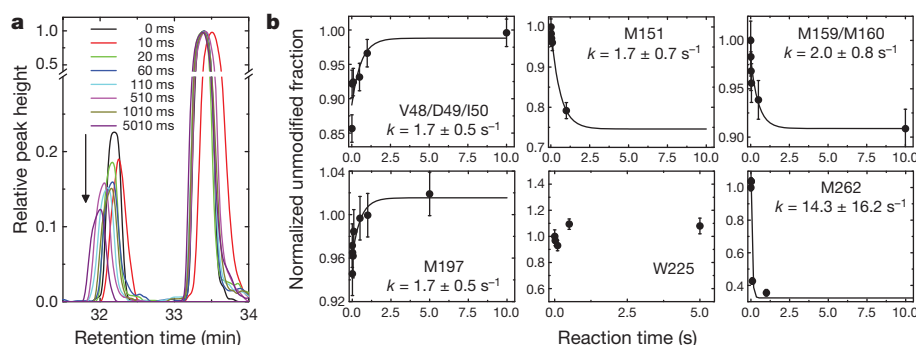


Figure 4 | Kinetics of water accessibility changes. **a**, An example of extracted ion chromatograms from the unmodified and modified M197 in peptide 197–208. The data were smoothed by a low-pass filter and normalized to the peak height of respective unmodified species. The arrow indicates a progressive decrease of the modified peaks as a function of the reaction time. **b**, Time

courses of water accessibility change for indicated residues. The solid line represents a single exponential fit of the time course of the unmodified fraction with a fitted rate constant (k) presented as mean \pm s.e.m. from six independent measurements.

(Extended Data Fig. 5). The side-chain-dependent effects of L152 mutations on protein stability are consistent with a critical structural role for L152 in the highly conserved TM5→TM3–TM6 packing core.

Among all the detectable sites, only the transport site and its neighbouring L152 gate exhibited all-or-nothing water accessibility changes (Figs 1d and 2a and Extended Data Fig. 3), suggesting a tight control of water leakage across the membrane. Within the TMD, oxidative modifications were observed at four reactive Met residues outside the transport site and the L152 gate. As noted above, M197 at the intracellular entrance to the L152 gate (Fig. 2b) showed a 70% reduction in water accessibility upon Zn(II) binding (Fig. 2a) whereas M151, M159 and M160 at the amino (N) and carboxy (C) termini of TM5 (Fig. 2b) showed a 50–130% increase in water accessibility (Fig. 2a and Extended Data Table 1). These last three Met residues reside on a helical face of TM5 with increased water accessibility upon Zn(II) binding (Fig. 2b). By contrast, residues with a reduction of water accessibility upon Zn(II) binding are either located on the opposite TM5 face (for example, L152) or packed against the opposite TM5 face (V48 and M197) (Fig. 2b). The tetrahedral transport site (H153, D157, D45 and D49) is also located on the same TM5 face with a zinc-dependent loss of water accessibility. The reciprocal change in water accessibility on two opposite TM5 faces is consistent with re-orientation of TM5 in response to Zn(II) binding. Furthermore, solvent-accessible residues in apo-YiiP were found to line a putative transmembrane zinc pathway, starting from M197 in the intracellular cavity, through L152 within the inter-cavity seal and arriving at H153, V48 and D49 in the extracellular cavity (Fig. 2b). This finding of a well-defined channel from the transport site in apo-YiiP to the intracellular cavity is in agreement with an inward-facing

conformation revealed by an electron crystallographic structure of an apo-YiiP homologue²¹.

To understand the structural dynamics of the Zn-dependent closing of the inter-cavity portal, we monitored the time course of radiolytic modification upon rapid mixing of apo-YiiP and 0.2 mM ZnCl₂ (Extended Data Fig. 1b). Only highly reactive residues could be detected with a sufficient signal-to-noise ratio for quantitative kinetic analysis (Fig. 4a). In the TMD, time-resolved measurements were performed on four Met residues (M151, M159/M160 and M197) and the V48/D49/I50 peptide. After mixing of apo-YiiP and Zn(II), the exponential increases in unmodified V48/D49/I50 and M197 residues, indicative of the closing of the inter-cavity portal, mirrored the exponential falls in unmodified M151 and M159/M160 residues (Fig. 4b). The rates of reciprocal water accessibility changes for these four positions on opposite faces of TM5 were identical within experimental errors, suggesting that TM5 underwent a rigid-body re-orientation upon zinc binding. The rigid-body motion of TM5 predicted a similar rate of L152 motion. Averaging the rates of four detectable sites gave an overall rate of TM5 motion at $1.8 \pm 0.7 \text{ s}^{-1}$, approximating the macroscopic transport rate ($2\text{--}5 \text{ s}^{-1}$) determined by stopped-flow Zn(II) flux measurements^{3,7}. Thus, zinc access to the transport site and the ensuing TM5 motion linked to the closing of the L152 gate occur on the same time scale. As an internal control, time-resolved measurement showed no changes in labelling to W225 on the cytoplasmic domain (CTD) surface (Fig. 4b). However, another surface residue, M262, exhibited a rapid change at 14.3 s^{-1} (Fig. 4b). The marked kinetic difference suggested that the observed water accessibility change to M262 preceded TM5 motion, but its functional relevance is unclear.

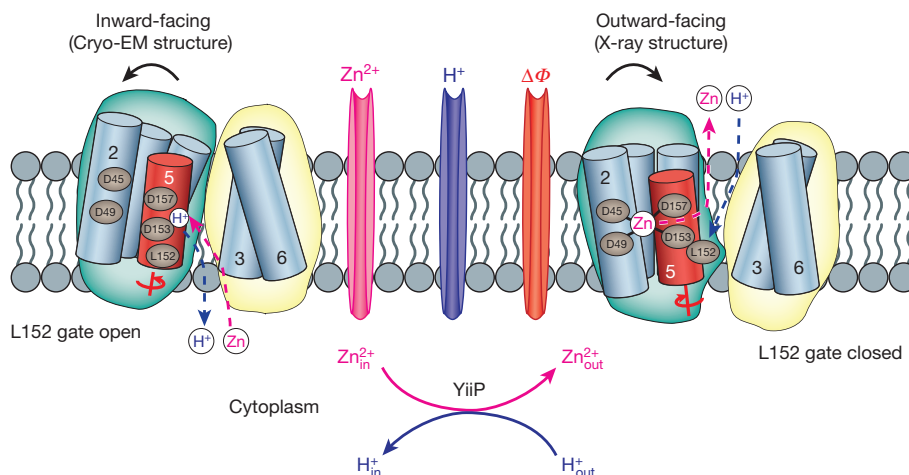


Figure 5 | Schematic representation of zinc-for-proton exchange. The representation is based on two existing structural models with the L152 gate open or closed as indicated. The protein conformational change alternates the

membrane-facing on–off mode of zinc coordination and protonation–deprotonation of the transport site in a coordinated fashion.

The time-resolved data suggested that Zn(II) binding triggers a concerted rigid-body motion of TM5 that swings L152 into place to plug the inter-cavity seal (Fig. 5). Since the four coordination residues of the transport site are projected from TM2 and TM5 (Fig. 3a), a TM5 motion is expected to alter the TM2–TM5 inter-helix orientation that determines the coordination geometry of the transport site¹. Thus, a rigid-body TM5 motion would simultaneously affect the mode of Zn(II) coordination and the gating of the inter-cavity portal through L152 movement (Fig. 5). As shown in Fig. 3c, d, the opening of the L152 gate would expose the transport site through a nanotube to the aqueous bulk of the intracellular cavity²². When a Zn(II) from the intracellular cavity reaches the transport site, the favourable match of its coordination chemistry with the tetrahedral transport site²³ would release binding free energy in the confinement of the hydrophobic core where the free energy may be guided to trigger TM5 re-orientation. By analogy to the working of a combustion engine, the released zinc-binding energy is transformed to useful mechanical energy, providing the power stroke of TM5 re-orientation to close the L152 gate (Fig. 5). As a result, this conformational change alternatively exposes the transport site to intracellular and extracellular cavity. The *in vivo* transmembrane proton gradient of an enteric bacterium *E. coli* is about one to two pH units²⁴. The flipping of H153 as a part of the transport site to either side of the membrane with a physiological pH gradient is expected to change its protonation state. A deprotonated H153 facing a relatively alkaline cytosol would promote Zn(II) binding from the intracellular cavity whereas a protonated H153 facing a relatively acidic periplasm may facilitate Zn(II) release into the extracellular cavity (Fig. 5). As such, an inward pH gradient drives a vectorial Zn(II) efflux in a 1:1 exchange stoichiometry. The dynamic details revealed in the present study explain how a physiological proton gradient, zinc coordination chemistry and water nanofluidics are orchestrated in a dynamic protein structure to overcome the activation barrier to Zn(II) efflux and promote a vectorial Zn(II) movement through an inter-cavity water portal that is highly conserved in the CDF protein family.

METHODS SUMMARY

YiiP was overexpressed and purified as described previously³. Prior to X-ray irradiation, YiiP was de-metalized and then exchanged to a radiolytic labelling buffer (10 mM NaPi, pH 6.5, 100 mM NaCl, 0.02% DDM, 0.1 mM TCEP) by size-exclusion HPLC. ZnCl₂ was added to 0.1 mM to an aliquot of apo-YiiP to form zinc-YiiP. Apo-YiiP or zinc-YiiP at a concentration of 10 μ M were exposed to an X-ray white beam at 4 °C at beamline X28C of the National Synchrotron Light Source, Brookhaven National Laboratory, as described previously^{16,25}. Proteolytic cleavage of the irradiated samples used pepsin, trypsin or trypsin–chymotrypsin double digestion. A bottom-up proteomic analysis by reverse phase liquid chromatography interfaced to a Fourier transform mass spectrometer. The MS/MS data for peptides and their sites of modifications were manually interpreted with the aid of proteomics software²⁶. The peak area from the extracted ion chromatograms of a specific peptide fragment was used to quantify the amount of modification. The extent of modification versus the X-ray irradiation time was fitted to a single exponential function to determine the hydroxyl radical reactivity rate of the side chain. Time-resolved radiolysis was performed with a modified KinTek apparatus using a standard flow sequence^{18,27}. Apo-YiiP at a concentration of 20 μ M, and an equal volume of 0.2 mM ZnCl₂ were mixed by a T-mixer. After a designated delay time, the mixed sample was driven through an irradiation cell and then into a quenching tube. The extent of radiolytic modification was plotted against the reaction time and fitted with a single exponential function to determine the rate of water accessibility change. All data are presented as mean \pm s.e.m. based on three or more independent measurements.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 October 2013; accepted 14 April 2014.

Published online 22 June 2014.

1. Lu, M., Chai, J. & Fu, D. Structural basis for autoregulation of the zinc transporter YiiP. *Nature Struct. Mol. Biol.* **16**, 1063–1067 (2009).
2. Grass, G. *et al.* FieF (YiiP) from *Escherichia coli* mediates decreased cellular accumulation of iron and relieves iron stress. *Arch. Microbiol.* **183**, 9–18 (2005).

3. Chao, Y. & Fu, D. Thermodynamic studies of the mechanism of metal binding to the *Escherichia coli* zinc transporter YiiP. *J. Biol. Chem.* **279**, 17173–17180 (2004).
4. Wei, Y. & Fu, D. Selective metal binding to a membrane-embedded aspartate in the *Escherichia coli* metal transporter YiiP (FieF). *J. Biol. Chem.* **280**, 33716–33724 (2005).
5. Wei, Y. & Fu, D. Binding and transport of metal ions at the dimer interface of the *Escherichia coli* metal transporter YiiP. *J. Biol. Chem.* **281**, 23492–23502 (2006).
6. Wei, Y., Li, H. & Fu, D. Oligomeric state of the *Escherichia coli* metal transporter YiiP. *J. Biol. Chem.* **279**, 39251–39259 (2004).
7. Chao, Y. & Fu, D. Kinetic study of the antiport mechanism of an *Escherichia coli* zinc transporter, ZitB. *J. Biol. Chem.* **279**, 12043–12050 (2004).
8. Kambe, T., Yamaguchi-Iwai, Y., Sasaki, R. & Nagao, M. Overview of mammalian zinc transporters. *Cell. Mol. Life Sci.* **61**, 49–68 (2004).
9. Montanini, B., Blaudez, D., Jeandroz, S., Sanders, D. & Chalot, M. Phylogenetic and functional analysis of the cation diffusion facilitator (CDF) family: improved signature and prediction of substrate specificity. *BMC Genomics* **8**, 107 (2007).
10. Nies, D. H. Efflux-mediated heavy metal resistance in prokaryotes. *FEMS Microbiol. Rev.* **27**, 313–339 (2003).
11. Palmiter, R. D., Cole, T. B., Quaife, C. J. & Findley, S. D. ZnT-3, a putative transporter of zinc into synaptic vesicles. *Proc. Natl Acad. Sci. USA* **93**, 14934–14939 (1996).
12. Lemaire, K. *et al.* Insulin crystallization depends on zinc transporter ZnT8 expression, but is not required for normal glucose homeostasis in mice. *Proc. Natl Acad. Sci. USA* **106**, 14872–14877 (2009).
13. Lu, M. & Fu, D. Structure of the zinc transporter YiiP. *Science* **317**, 1746–1748 (2007).
14. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543 (2007).
15. Takamoto, K. & Chance, M. R. Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 251–276 (2006).
16. Gupta, S. *et al.* Conformational changes during the gating of a potassium channel revealed by structural mass spectrometry. *Structure* **18**, 839–846 (2010).
17. Angel, T. E., Gupta, S., Jastrzebska, B., Palczewski, K. & Chance, M. R. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl Acad. Sci. USA* **106**, 14367–14372 (2009).
18. Gupta, S., D'Mello, R. & Chance, M. R. Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. *Proc. Natl Acad. Sci. USA* **109**, 14882–14887 (2012).
19. Podar, D. *et al.* Metal selectivity determinants in a family of transition metal transporters. *J. Biol. Chem.* **287**, 3185–3196 (2012).
20. Lin, H. *et al.* Gain-of-function mutations identify amino acids within transmembrane domains of the yeast vacuolar transporter Zrc1 that determine metal specificity. *Biochem. J.* **422**, 273–283 (2009).
21. Coudray, N. *et al.* Inward-facing conformation of the zinc transporter YiiP revealed by cryoelectron microscopy. *Proc. Natl Acad. Sci. USA* **110**, 2140–2145 (2013).
22. Hinds, B. J. *et al.* Aligned multiwalled carbon nanotube membranes. *Science* **303**, 62–65 (2004).
23. Hoch, E. *et al.* Histidine pairing at the metal transport site of mammalian ZnT transporters controls Zn²⁺ over Cd²⁺ selectivity. *Proc. Natl Acad. Sci. USA* **109**, 7202–7207 (2012).
24. Ramos, S., Schuldiner, S. & Kaback, H. R. The electrochemical gradient of protons and its relationship to active transport in *Escherichia coli* membrane vesicles. *Proc. Natl Acad. Sci. USA* **73**, 1892–1896 (1976).
25. Gupta, S., Sullivan, M., Toomey, J., Kiselar, J. & Chance, M. R. The Beamline X28C of the Center for Synchrotron Biosciences: a national resource for biomolecular structure and dynamics experiments using synchrotron footprinting. *J. Synchrotron Radiat.* **14**, 233–243 (2007).
26. Kaur, P., Kiselar, J. G. & Chance, M. R. Integrated algorithms for high-throughput examination of covalently labeled biomolecules by structural mass spectrometry. *Anal. Chem.* **81**, 8141–8149 (2009).
27. Ralston, C. Y. *et al.* Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Methods Enzymol.* **317**, 353–368 (2000).

Acknowledgements We dedicate this work to the memory of Peter C. Maloney, who read and commented on a version of the manuscript. This work was supported in part by the National Institutes of Health under grant R01 GM065137 (to D.F.), the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the US Department of Energy (DOE) under contract DE-AC02-98CH10886 (to D.F.); D.F. is primarily supported by the Physical Biosciences Program, Office of Basic Energy Sciences of the DOE and the National Institute for Biomedical Imaging and Bioengineering under grants P30-EB-09998 and R01-EB-09688 (to M.R.C.). The National Synchrotron Light Source at Brookhaven National Laboratory is supported by the DOE under contract DE-AC02-98CH10886.

Author Contributions M.R.C. and D.F. conceived the work, S.G. and D.F. designed the experiments, S.G., J.C., R.D. and D.F. performed the experiments, S.G., M.R.C. and D.F. analysed the data, and D.F. interpreted the data and wrote the paper with S.G. and M.R.C.

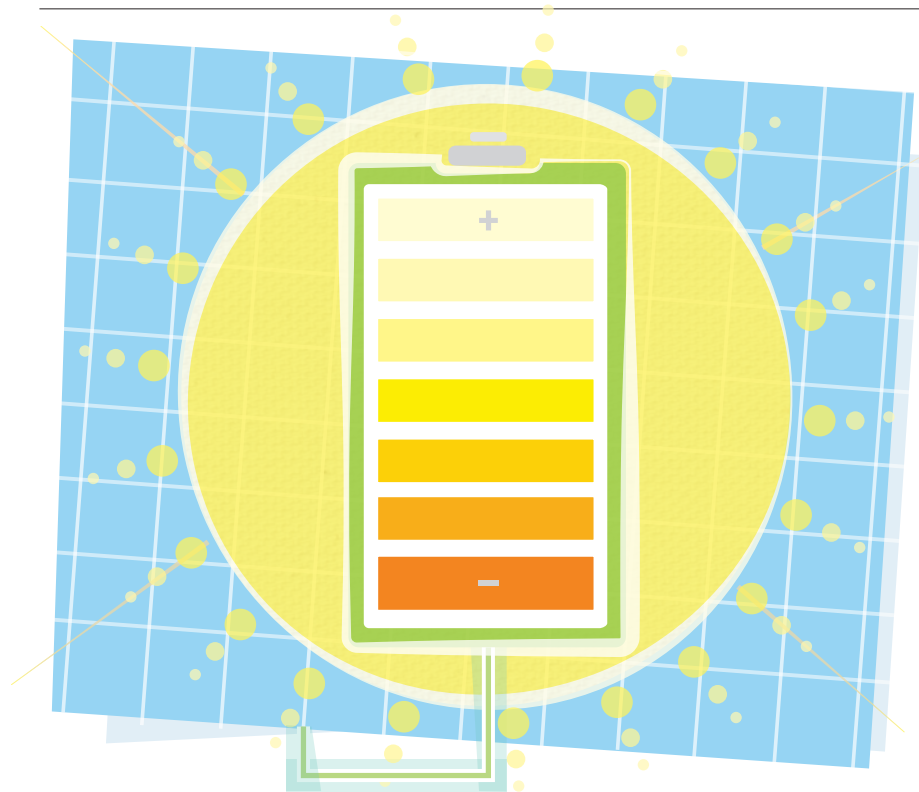
Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.F. (dfu3@jhmi.edu) or M.R.C. (mrc16@case.edu).

CAREERS

OUTSOURCING US drug-makers shift clinical trials to contractors **p.106**

EMPLOYMENT LAW Association argues for US graduate assistants' right to unionize **p.106**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



MATERIALS RESEARCH

Batteries warm up

Interest in energy-storage research is growing, opening up opportunities for chemists with interdisciplinary skills.

BY KATHARINE GAMMON

For materials scientist Lynn Trahey, the impetus for a change in career direction was a single sentence she heard in 2005. Then a doctoral student in chemistry at the University of California, Berkeley, she was attending a talk about the potential of powering the world with solar energy. The speaker made an offhand comment — that energy storage is part of the equation for the future of green energy — and Trahey's interest was piqued. "I took that and ran with it — he's right. We will need to store energy from all kinds of sources. Batteries have to be part of the innovation," she says.

Trahey completed her studies in 2007 and went hunting for a postdoctoral position that would satisfy her interest in lithium-ion

batteries, the most common rechargeable battery used in electronics and cars. She landed a joint postdoc in 2008 at two institutes in Illinois — Argonne National Laboratory near Chicago and Northwestern University in Evanston — then nabbed a permanent position as an assistant materials scientist at the lab two years later.

Today, she is researching materials and reactions in high-energy-density rechargeable cells with the aim of making longer-lasting batteries. Specifically, she wants to learn why the devices degrade where their liquid and solid components connect. She is also working to develop anodes — where positive current flows into the battery — from materials such as tin and silicon, which are cheaper and last longer than lithium.

The world is gradually moving to greener sources of energy, but trapping that power is

troublesome because energy is lost every time it is moved or converted. Governments and industry are investing heavily to improve energy storage, and this is translating into research opportunities for early-career scientists who have skills in chemistry, electrochemistry and materials science, and are familiar enough with physics and engineering to discuss energy storage with physicists and engineers.

Trahey recommends that doctoral students who are interested in the field get a multi-disciplinary education, particularly in the fields of electrochemistry and chemical engineering. In her battery group, she is surrounded by materials engineers and physicists who communicate well with one another. "We speak different languages on the same topic," she says, but she learned on the job how to make sure she understands — and is understood by — those in other specialties. She advises early-career researchers to hone these skills by attending talks on energy storage that lie outside their specific field.

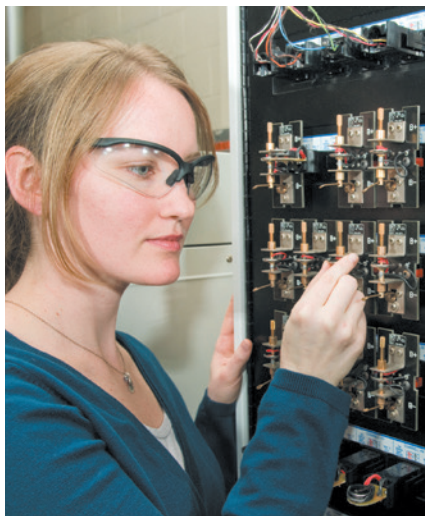
BOUNDARY BREAKERS

A collaborative mindset helps to boost employability because the battery field requires a particularly cooperative spirit owing to its complexity. Devin Hodge oversees hiring for the Joint Center for Energy Storage Research (JCESR) in A, a five-year, US\$120-million partnership between government, industry and academia that is funded by the US Department of Energy (DOE) and located at Argonne. He says that scientists who want to work effectively in an energy-storage laboratory should have a spirit of innovation as well as collaboration.

Hodge says that research opportunities will be opening up at the JCESR in the next year or so, but not all is rosy in the field's hiring outlook. Funding, at least in US academia, could well tighten; although federal spending on battery research has risen in the past five years, researchers hoping to grab a slice of that pie have glutted the field to some extent, says Brent Melot, a chemist at the University of Southern California in Los Angeles. "Some people say that it's the worst funding environment because of the number of people who are now competing for the same opportunities," he says. "All the people who used to research magnets now work on energy storage."

But industry, including the car and electronic-device sectors, is not stymied by the same funding crunch as academia. And the field of energy storage is growing around the world — meaning more jobs. Globally, government-funded battery and fuel-cell research and development ►

CLAIRE WELSH/NATURE



Lynn Trahey is researching anode materials.

► accounted for \$8.7 billion from 2008 to 2012, and that number is expected to increase. The French government has allotted €140 million (US\$189 million) to automotive-battery research. And energy-storage associations also cropped up in Asia: the Korea Battery Industry Association started in 2011; India's version launched in 2012; and China's Energy Storage Alliance started in 2010. Each of these associations supports research efforts and helps to move information to partners who can use it.

The JCESR — launched by the DOE in 2012 — now has 14 national partners, including Johnson Controls in Glendale, Wisconsin, and the Dow Chemical Company in Midland, Michigan. It aims to develop batteries that generate power in different ways, partnering lithium with oxygen or sulphur, for example (see *Nature* 507, 26–28; 2014). That means opportunities for early-career scientists in chemistry, electrochemistry, materials science, nanotechnology, chemical engineering, computation and mechanical engineering.

In the next wave of hiring, some of the current postdocs will move into staff-scientist positions internally. Hodge says that graduate students, and postdocs in particular, will learn skills such as computational and materials techniques, which will effectively position them for high-paying permanent positions at research centres or in industry.

STORAGE STARS

A constellation of battery-related start-up businesses has emerged across the United States, and prospects abound for young, talented researchers who can innovate using chemistry. One of those is Vincent Giordani, a senior scientist at Liox Power, an energy research-and-development company in Pasadena, California.

Giordani acknowledges that his career path was serendipitous. When he presented his doctoral thesis on rechargeable non-aqueous lithium-air batteries at a conference in

Canada, two executives from Liox happened to be in the audience. He was hired soon after, and moved to the United States two weeks after finishing his PhD jointly in France and the United Kingdom in 2010.

Industry presents interesting challenges for scientists in battery research, he says. "In academia, you have to bring money and write papers, but in industry there are higher stakes in coming up with a new technology." He is working on batteries that are more recyclable, hold more energy and can push cars farther.

Indeed, the automotive industry is focusing on creating cheaper and longer-lasting batteries, which should produce opportunities for early-career researchers. "The auto industry will change more in the next 10 years than it has in the past 100 years — a battery is no longer a commodity, but an integral part of the vehicle itself, with increasing demands for power and longevity," says David Cue, a vice-president at Johnson Controls Power Solutions in Milwaukee, Wisconsin, which is hiring PhD graduates with a background in engineering and materials.

To succeed in battery research, it helps to have a broad view of the necessary components. The shift from characterizing materials to working in a battery system can be daunting. "It's especially challenging for chemists because energy storage is an engineering-focused world," says Melot. "If you're not able to focus on the whole device, you go down dead-end roads." He recommends that early-career researchers get immersed in the wider field by gaining experience in labs that look at a whole device instead of one tiny part.

Despite academia's uncertain outlook, the interest in — and funding for — a world of

"The auto industry will change more in the next 10 years than it has in the past 100 years."

better batteries continues to rise.

In late June, the DOE announced a \$3.2-million investment in the National Incubator Initiative for Clean

Energy, which will create a network to assist small businesses focused on clean energy in honing their ideas and bringing green products to market. The incubators should help small businesses to grow, creating more jobs for early-career researchers.

The International Energy Agency estimates that by 2018 one-quarter of the electricity produced worldwide will come from renewable sources. Wherever clean energy is, there will always be a need to store it in a better way. "There are a lot of opportunities," says Cue, who has worked in Germany, China and France as well as in the United States. "And there aren't enough electrochemists to fill the demand." ■

Katharine Gammon is a freelance writer in Santa Monica, California.

RESEARCH AND DEVELOPMENT

Outsourcing trends

US drug-makers are outsourcing more and more of their research and development, mainly to contract research organizations, according to figures released in July by the US National Science Foundation (NSF). The shift could be good news for researchers seeking positions in industry. In 1991, pharmaceutical companies spent about US\$800 million on external research and development, but that skyrocketed to \$13 billion in 2011, says John Jankowski, head of research and development statistics at the NSF. That growth outstrips that of any other sector. In 1991, industrial extramural research spending totalled \$3.3 billion, but by 2011, spending had risen to \$25.3 billion for domestic companies alone. Pharmaceutical firms' share of that total was 23% in 1991, but ballooned to 51% by 2011. Jankowski says that much of the increase comes from the outsourcing of clinical trials. The number of US contract research organizations has risen to match the demand, from around 800 in 2000 to more than 3,100 by the end of 2011, according to the Tufts Center for the Study of Drug Development in Boston, Massachusetts.

EMPLOYMENT LAW

Graduate rights

The American Association of University Professors in Washington DC has filed a legal document to the National Labor Relations Board arguing that graduate assistants, including research technicians, at private institutions should be considered employees and should therefore have collective-bargaining rights. The brief argues that the board should revise its definition of employee status, which is based on a 2004 decision that graduate assistants at Brown University in Providence, Rhode Island, were not employees because their work was inextricably linked to their study. Union representation of graduate assistants is a contentious issue. In 2012, Michigan banned graduate-student research assistants in public universities from unionizing, arguing that giving students employee status would alter the student-teacher relationship. In 2008, research assistants at the Research Foundation of the State University of New York voted to elect union representation after a 2007 board ruling that they were fundamentally employees.

YOUR APPLICATION FOR ETERNAL LIFE HAS BEEN PARTIALLY APPROVED

All that matters.

BY JAMES WESLEY ROGERS

This is an important message from the Central Existence Committee. Press 1 to receive this message in English. Press 2 to receive this message in Spanish. Press 3 to have this message directly injected into your brain using Chomsky universal grammar. You have three seconds to choose.

Congratulations, Mr Lawson. We have confirmed that you have less than 6 weeks to live. You are now eligible for total fermionic regeneration. Your application has been approved with a continuity coefficient of 0.8 (80%).

You will be regenerated via a process originally developed for quantum teleportation. First, you will be subatomically scanned, using what is referred to as a 'destructive read'. Essentially, your body will be completely destroyed as it is converted into pure information. Next, we will apply powerful error-correcting algorithms to your data, and reconstitute your body with all pathological conditions eliminated. Your data will not be shared with any third party.

During reconstitution, your personality will be slightly altered, using a simple linear interpolation between yourself (80%) and a standard personality template (20%).

Your continuity coefficient of 0.8 (80%) was determined by a number of factors:

1) Your current profession, astronaut, is considered nonessential. By adjusting your aptitudes, we hope to nudge you towards a more useful vocation, such as obesity counselling or motivational performance art.

2) You have a poor record of separating your recyclables. To have any chance at all of being an effective obesity counsellor, you will need a greater sense of civic responsibility.

3) We have detected a disturbing pattern of negativity towards Canadian pop stars on your social media accounts. Such cynicism and poor aesthetic judgement have no place in an enlightened society.

As your continuity coefficient is less than the legal threshold of 94% established by *Jones-B v. California*, your derived self will not legally be Linus Lawson, but rather your direct descendant, Linus Lawson-B.

➔ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm

You may transfer up to 100% of your current assets to Linus

Lawson-B without probate or inheritance tax. Your academic and professional qualifications will also be transferred, contingent on a series of review examinations.

Linus Lawson-B's age will be legally established as 21 at the time of regeneration. Any contractual agreement, including marriage, to which Linus Lawson is a party may be voided at this time.

Press 1 to review frequently asked questions about fermionic regeneration. Press 2 to proceed directly to the next section. You have three seconds to choose.

Frequently asked questions Will the reconstituted person really be me?

For a continuity coefficient of 1.0 (100%), the general consensus is yes, both legally and ontologically. For lower continuity coefficients, there is much debate.

Ask yourself this: am I the same person when I wake up in the morning as I was the night before? Research has shown that during periods of intense synaptic reorganization, the continuity coefficient of someone waking from a deep sleep can be as low as 0.996 (99.6%) As neural adaptation and memory degradation accumulate every day, are you still the same person you were one year ago?

It has been postulated that a human being is not one continuous person throughout his or her life, but rather a sequence of discrete persons, each with merely a perception of continuity with the preceding individuals. Perhaps it is this perception of continuity that is the functional definition of a soul. If that is the case, and your derived self has this perception of continuity with you, then he or she will be you in the only meaningful sense.

The most we can say then, is that there is a very high probability that your derived self will have a perception of continuity with you.

What about my loved ones?

All memories of your friends and family members will be transferred, and your

derivative self may retain some of your emotional attachment to them.



Will I ever need to be regenerated again?

Barring accident, no. Genetic and cellular improvements made during error correction will result in a derivative body that can be indefinitely maintained in optimal condition by conventional therapies.

Will you keep a back-up of my data?

Unfortunately, this is a practical impossibility. The amount of information needed to encode a human being at the subatomic level is unimaginably huge. At this time, we possess only enough information-storage capacity to tempo-

rarily encode a single human being, and to keep one standard template. The standard personality template was chosen from among the giants of human intellect and accomplishment of the past seven decades.

I've heard that the standard personality template is legendary Canadian pop star Justin Bieber. Is that true?

We can neither confirm nor deny this rumour. End of FAQ

Once again, Mr Lawson, congratulations. You are about to embark on a great adventure. Due to scheduling constraints, your regeneration must begin immediately. Should you choose to decline regeneration at this time, you will not be allowed to reapply. Press 1 to be completely regenerated into a functionally immortal form that will retain 80% of your current identity. Press 2 to decline regeneration and live for up to 6 more weeks as Linus Lawson. You have three seconds to choose. ■

James Wesley Rogers sometimes allows computational geometry to distract him from science fiction. These distractions occur most often at a 3D modelling software company in Columbus, Ohio.

JACEY